ENSEMBLE-BASED DATA ASSIMILATION:

A REVIEW

Thomas M. Hamill

University of Colorado and NOAA-CIRES Climate Diagnostics Center

Boulder, Colorado, USA

29 July 2003

Corresponding author address: Dr. Thomas M. Hamill, NOAA-CIRES Climate Diagnostics Center, R/CDC 1, 325 Broadway, Boulder, Colorado USA 80305-3328. tom.hamill@noaa.gov ; 1 (303) 497-3060 ; telefax 1 (303) 497-7013

ABSTRACT

The literature on ensemble-based data assimilation techniques has been growing rapidly in past decade. These techniques are being explored as possible alternatives to current operational analysis techniques such as 3- or 4-dimensional variational assimilation. Ensemble-based assimilation techniques utilize an ensemble of parallel data assimilation and forecast cycles. The background-error covariances are estimated using the forecast ensemble and are used to produce an ensemble of analyses. The background-error covariances are flow dependent and often have very complicated structure, providing a very different adjustment to the observations than are seen from methods such as 3- dimensional variational assimilation. Though computationally expensive, ensemble-based techniques are relatively easy to code, since no adjoint nor tangent-linear models are required, and tests in simple models suggest that dramatic improvements over existing operational methods may be possible.

A review of the ensemble-based assimilation is provided here, starting from the basic concepts of Bayesian assimilation. Without some approximation, Bayesian assimilation is computationally impossible for large-dimensional systems. Assuming normality of error statistics and linearity of error growth, the state and its error covariance may be predicted optimally using Kalman filter techniques. Explanations of ensemble assimilation methods are then provided. If linearity and normality assumptions are valid, as ensemble size increases, the mean and covariance estimates from ensemble assimilation methods will converge to those produced by the extended Kalman filter. However, for high-dimensional states, ensemble assimilation methods are computationally less expensive than Kalman filters and may outperform them when linear and normal assumptions are inappropriate. The accuracy of both methods depend on reasonable models of error statistics.

Techniques for making the ensemble assimilation methods more accurate and more computa-

tionally efficient on parallel computers are discussed, and an example of ensemble data assimilation of a sparse network of real surface pressure observations into a global numerical weather prediction model is provided.

1. INTRODUCTION

The purpose of this article is to introduce the reader to promising new experimental methods for atmospheric data assimilation involving the use of ensemble forecasts (e.g., Evensen 1994, Evensen and van Leeuwen 1996, Houtekamer and Mitchell 1998, 1999, 2001, Burgers et al. 1998, van Leeuwen 1999, Lermusiaux and Robinson 1999, Anderson and Anderson 1999, Miller et al. 1999, Hamill and Snyder 2000, 2002, Keppenne 2000, Mitchell and Houtekamer 2000, Heemink et al. 2001, Hamill et al. 2001, 2003, Anderson 2001, Pham 2001, Verlaan and Heemink 2001, Bishop et al. 2001, Keppenne and Rienecker 2002, Whitaker and Hamill 2002, Mitchell et al. 2002, Hansen 2002, Lermusiaux 2002, Reichle et al. 2002ab, Reichle and Koster 2003, Snyder and Zhang 2003, Tippett et al. 2003, Anderson 2003, Zhang et al. 2003, Evensen 2003, Lawson and Hansen 2003, Ott et al. 2003). There is a natural linkage between data assimilation and ensemble forecasting. Ensemble forecasts (Toth and Kalnay 1993, 1997, Molteni et al. 1996, Houtekamer et al. 1996a) are designed to estimate the flow-dependent uncertainty of the forecast; data assimilation techniques require accurate estimates of forecast uncertainty in order to optimally blend the prior forecast(s) with new observations. Ensemble-based assimilation methods integrate the two steps; the ensemble of forecasts is used to estimate forecast-error statistics during the data assimilation step, and the output of the assimilation is a set of analyses. This process is cycled, the short-term ensemble forecasts from the set of analyses providing the error statistics again for the next assimilation cycle.

Four-dimensional variational analysis (4D-Var; Le Dimet and Talagrand 1986, Courtier et al. 1994, Rabier et al. 1998, 2000) is now considered the state-of-the-art technique for atmospheric data assimilation. 4D-Var finds the model trajectory that best fits the observational data over a period of time. It is thus worth questioning why it may be worthwhile to consider such a different technique. Will ensemble-based assimilation methods produce more accurate analyses? This cannot be answered definitively yet. Direct comparisons of ensemble assimilation methods and 4D-Var in realistic scenarios have yet to be performed. Right now 4D-Var is relatively well established, while ensemble-based methods have demonstrated dramatically reduced

errors in controlled tests with simple models. Testing of ensemble-based methods in more complex models with real data has only been started within the last few years with suggestive but inconclusive results.

Ensemble-based methods have some potential advantages that make them worthy of further consideration. Ensemble-based methods are much easier to code and maintain, for neither a tangent linear nor an adjoint model of the forecast model is required, as they are with 4D-Var. Ensemble-based methods produce an ensemble of possible analysis states, providing information on both the mean analysis and its uncertainty. In comparison, 4D-Var estimates only the mean. This has two useful consequences for ensemble-based methods. First, an ensemble of initial conditions is available for subsequent ensemble forecasts; no additional computations are required to create perturbed initial conditions, as are required with 4D-Var. Second, if the analysis uncertainty is very spatially inhomogeneous and time dependent, in ensemble-based methods this information will be fed through from one assimilation cycle to the next. In comparison, in 4D-Var, the assimilation typically starts at each update cycle with the same stationary model of error statistics. Hence, the influence of observations may be more properly weighted in ensemble-based methods than in 4D-Var.

Another potential strength of ensemble-based methods is their ability to incorporate the effects of forecast uncertainty that are due to model imperfections directly into the data assimilation. In comparison, in current operational implementations of 4D-Var, the forecast model dynamics are a strong constraint (Courtier et al. 1994; but see Bennett et al. 1996 for a possible alternative). If the forecast model used in 4D-Var does not adequately represent the true dynamics of the atmosphere, model error may be large, and 4D-Var may fit a model trajectory that was significantly different than the trajectory of the real atmosphere during that time window.

Ensemble-based techniques may have disadvantages relative to 4D-Var, some that will only be discovered through experimentation. Ensemble-based techniques are likely to be at least as computationally expensive as 4D-Var, and early tests showed that without careful adjustment, the method is not robust and will underestimate forecast errors and insufficiently draw to new

observational data. Further, 4D-Var is more naturally suited to assimilating observations that are uniformly distributed in time such as satellite observations, and it may handle observations with nonlinear observation operators such as satellite radiances better than ensemble techniques (Lorenc 2003). Ensemble approaches may be difficult to apply in limited-area models because of the lateral boundary conditions, and balance issues may be problematic (ibid). In the end, it will require significant further research to determine whether ensemble assimilation methods can outperform 4D-Var. Given the advantages cited above, however, there is reason for optimism and further testing. This review will provide a reference for ensemble-based assimilation and hopefully will stimulate wider interest in the technique.

Rather than starting with the specifics of recently proposed ensemble-based assimilation techniques, in this paper we will take a step back and try to motivate their use by quickly developing them from first principles, noting the approximations that have been made along the way. This will take us from Bayesian data assimilation (section 2), which is conceptually simple but computationally prohibitive, to the Kalman filter (section 3), a simplification assuming normality and linearity of error growth, to ensemble assimilation methods (section 4), which may be more computationally tractable and robust. We will then discuss some of the algorithmic techniques (sections 5-6) and demonstrate an ensemble-based assimilation method using a sparse network of surface pressure observations and a global numerical weather prediction model (section 7).

Another contemporaneous review paper on ensemble-based data assimilation is available (Evensen 2003). This paper provides less background material on Bayesian assimilation and the roots of the Kalman filter, but it provides a wider review of the currently discussed ensemble-based assimilation approaches, a more theoretical examination of the treatment of model errors, and a wide array of references to ensemble-based assimilation in the oceanographic literature. Lorenc (2003) also reviews ensemble methods, and in particular provides some thoughts on the potential relative strengths and weaknesses compared to 4D-Var.

In subsequent discussion, the atmosphere state, which is of course a continuum, is assumed to be adequately described in discretized fashion, perhaps by the values of winds, temperature, humidity, and pressure at a set of grid points.

2. BAYESIAN DATA ASSIMILATION

Conceptually, the atmospheric data assimilation problem is a relatively simple one. The task at hand is to accurately estimate the probability density function (pdf) for the current atmospheric state given all current and past observations. Much of the material in this section follows the work of Jazwinski (1970). If the reader is interested in further material on the subject, Lorenc (1986) provides a formulation of data assimilation in a Bayesian context, and Talagrand (1997) and Daley (1997) provide excellent reviews of atmospheric data assimilation and in particular the Kalman filter discussed in Section 3. Cohn (1997) provides a more rigorous statistical formulation of the problem.

When considering Bayesian assimilation, there are two general steps to the assimilation. Assume that a pdf of the state of the atmosphere is available (in the lack of any knowledge, this may be the climatological pdf). The first step is to assimilate recent observations, thereby sharpening the pdf. The second step is to propagate the pdf forward in time until new observations are available. If the pdf is initially sharp (i.e., the distribution is relatively narrow), then chaotic dynamics and model uncertainty will usually broaden the probability distribution. These two steps are then repeated. We will describe each of these steps separately, starting with the assimilation of new observations.

a. Bayesian updating

Assume that an estimate of the pdf has been propagated forward to a time when observations are available, then the state can be estimated more specifically by incorporating information from the new observations. This will be termed the "update."

The following notational convention is used. Boldface characters will denote vectors or matrices, while use of the italicized font denotes a scalar. $\mathbf{x}^{\mathrm{t}}_{\mathrm{t}-1}$ will denote the n-dimensional true

model state at time $t-1$: $\mathbf{x}^t_{t-1} = [x^t_{t-1(1)}, \ldots, x^t_{t-1(n)}]$. Also, assume a collection of observations ψ_t. This vector includes observations $\mathbf{y}_t$ at the most recent time as well as observations at all previous times $\psi_t = [\mathbf{y}_t, \psi_{t-1}]$, where $\psi_{t-1} = [\mathbf{y}_{t-1}, \ldots, \mathbf{y}_0]$. There are M_t observations at time t, i.e., $\mathbf{y}_t = [y_{t(1)}, \ldots, y_{t(M_t)}]$. Let $P(\mathbf{x}^t_t)$ be a multivariate probability density function, defined such that $Pr(\mathbf{a} \leq \mathbf{x}^t_t \leq \mathbf{b}) = \int_{\mathbf{a}}^{\mathbf{b}} P(\mathbf{x}^t_t)\, d\mathbf{x}^t_t$, and probability density integrates to 1.0 over the entire phase space.

Formally, the update problem is to accurately estimate $P(\mathbf{x}^t_t \mid \psi_t)$, the probability density estimate of the current atmospheric state given the current and past observations. Bayes' Rule tells us that this quantity can be re-expressed as

$$P(\mathbf{x}^t_t \mid \psi_t) \propto P(\psi_t \mid \mathbf{x}^t_t)\, P(\mathbf{x}^t_t). \tag{1}$$

Bayes' Rule is usually expressed with a normalization constant in the denominator on the right-hand side of (1); for simplicity, the term in the denominator will be dropped here, and it is assumed that when coded, the developer will ensure that probability density integrates to 1.0.

One hopefully minor assumption is made: observation errors are independent from one time to the next. Hence, $P(\psi_t \mid \mathbf{x}^t_t) = P(\mathbf{y}_t \mid \mathbf{x}^t_t)\, P(\psi_{t-1} \mid \mathbf{x}^t_t)$. This may not be true for observations from satellites, where instrumentation biases may be difficult to remove. Also, errors of observation representativeness (Daley 1991, 1993) may be flow dependent and correlated in time. But under this assumption, (1) is equivalent to

$$P(\mathbf{x}^t_t \mid \psi_t) \propto P(\mathbf{y}_t \mid \mathbf{x}^t_t)\, P(\psi_{t-1} \mid \mathbf{x}^t_t)\, P(\mathbf{x}^t_t). \tag{2}$$

By Bayes' Rule again, $P(\psi_{t-1} \mid \mathbf{x}^t_t)\, P(\mathbf{x}^t_t) \propto P(\mathbf{x}^t_t \mid \psi_{t-1})$. Hence, (2) simplifies to

$$P(\mathbf{x}^t_t \mid \psi_t) \propto P(\mathbf{y}_t \mid \mathbf{x}^t_t)\, P(\mathbf{x}^t_t \mid \psi_{t-1}). \tag{3}$$

In principle, equation (3) is elegantly simple. It expresses a recursive relationship: the "posterior," the pdf for the current model state given all the observations, is a product of the the probability distribution for the current observations $P(\mathbf{y}_t \mid \mathbf{x}^t_t)$ and the "prior," $P(\mathbf{x}^t_t \mid \psi_{t-1})$, also known

as the "background." The prior is the pdf of the model state at time t given all the past observations up to time $t-1$. Typically, the prior will have been estimated in some fashion from a cycle of previous data assimilations and short-term forecasts up to the current time; approximations of how this may be computed will be discussed in section 2b.

The expression $P(\mathbf{y}_t \mid \mathbf{x}_t^t)$ may be confusing: the observations are known, so why should this be expressed probabilistically? Consider for a moment that $\mathbf{y}_t$ is a random variable consisting of the converted known true state plus unknown random error: $\mathbf{y}_t = \mathcal{H}(\mathbf{x}_t^t) + \epsilon$, where $\mathcal{H}(\cdot)$ is a "forward" observation operator that converts the model state to the observation type and location. Let $\langle \cdot \rangle$ denote the expected value. If ϵ is normally distributed with $\langle \epsilon \rangle = 0.0$ and $\langle \epsilon \epsilon^{\mathrm{T}} \rangle = \mathbf{R}$, then

$$P(\mathbf{y}_t \mid \mathbf{x}_t^t) \sim exp\Big[-\frac{1}{2}\Big(\mathcal{H}(\mathbf{x}_t^t) - \mathbf{y}_t\Big)^{\mathrm{T}} \mathbf{R}^{-1}\Big(\mathcal{H}(\mathbf{x}_t^t) - \mathbf{y}_t\Big)\Big], \tag{4}$$

that is, $P\Big(\mathbf{y}_t \mid \mathbf{x}_t^t\Big) = P\Big(\mathbf{y}_t \mid \mathcal{H}(\mathbf{x}_t^t)\Big) \sim N\Big(\mathcal{H}(\mathbf{x}_t^t), \mathbf{R}\Big)$, the distribution is normal with mean $\mathcal{H}(\mathbf{x}_t^t)$ and covariance $\mathbf{R}$ (Cohn 1997). Of course, the observations are known and the truth is the random variable. In such a case, it is useful to think of $P\Big(\mathbf{y}_t \mid \mathcal{H}(\mathbf{x}_t^t)\Big)$ as equivalent to its likelihood $L\Big(\mathcal{H}(\mathbf{x}_t^t) \mid \mathbf{y}_t\Big)$, expressing the likelihood that the interpolated truth has the value $\mathcal{H}(\mathbf{x}_t^t)$ given the set of observations $\mathbf{y}_t$ (Casella and Berger 1990, p. 265). In (4), the probability density assigned to a particular $\mathcal{H}(\mathbf{x}_t^t) - \mathbf{y}_t$ difference is identical regardless of whether $\mathcal{H}(\mathbf{x}_t^t)$ or $\mathbf{y}_t$ is the random variable. Hence, the $P(\mathbf{y}_t \mid \mathbf{x}_t^t)$ in (3) can still be described by (4), but for purposes of Bayesian data assimilation, it is convenient to think of this as $P(\mathbf{y}_t \mid \mathbf{x}_t^t) \sim N(\mathbf{y}_t, \mathbf{R})$, a normal distribution centered on the actual observations.

Let's now demonstrate the update step of Bayesian assimilation with a simple example. $P(\mathbf{x}_t^t \mid \psi_{t-1})$ is an estimate of the prior for a two-dimensional model state. This was produced by assimilating all prior observations up to and including time $t-1$ and estimating in some manner how that pdf has evolved in the time interval between $t-1$ and t. Consider how to update the pdf given a new scalar observation y, which in this example is observing the same quantity as the first component of the state vector measures. The pdf for the observation $P(y_t \mid \mathbf{x}_t^t)$ is as-

sumed to be distributed normally about the actual observation, $\sim N(y_t, \sigma^2)$. Here, let $y_t = 58$ and $\sigma^2 = 100$.

Selected contours of the prior are plotted in Fig. 1a ; as shown, the prior is bimodal. The shape of the marginal prior distributions $P(x_{t(1)} \mid \psi_{t-1})$ and $P(x_{t(2)} \mid \psi_{t-1})$ are plotted along each axis in solid lines. The dashed line denotes the observation probability distribution $P(y_t \mid \mathbf{x}_t^t)$. This probability varies with the value $x_{t(1)}$, but given $x_{t(1)}$ is the same for any value of $x_{t(2)}$. The updated posterior distribution is computed using (3) and is shown in Fig. 1b. Note that the assimilation of the observation enhanced the probability in the lobe overlapping the observation distribution and decreased it in the other lobe. Overall, the posterior distribution is more sharp (specific) than the prior, as is expected.

b. Forecasting of probability density

The question of how the pdf estimate is propagated forward in time when new observations are available is now considered. Let $\mathcal{M}$ denote the discrete forward propagator, i.e., the deterministic weather forecast model. Ideally, the forecast model would be perfect, so that if the true state is known at time $t-1$, the future state at time t and all subsequent times can be defined: $\mathbf{x}_t^t = \mathcal{M}(\mathbf{x}_{t-1}^t)$. Unfortunately, weather prediction models have errors, a consequence of the finite resolution of the model and the resultant lack of interaction with unresolved scales as well as deficiencies in the model formulation. Consequently, a more realistic model of the time evolution is

$$\mathbf{x}_{t+1}^t = \mathcal{M}(\mathbf{x}_t^t) + \epsilon_{(t,t+1)}^x \tag{5}$$

where $\epsilon_{(t,t+1)}^x$ represents an (unknown) n-dimensional error of the model state incurred between time t and $t+1$. For ease of subsequent analysis, it is often assumed that the error can be modeled with a stochastic process:

$$\epsilon_{(t,t+1)}^x = \mathbf{G}_t(\mathbf{x}_t^t)\, \delta_t. \tag{6}$$

Here, $\mathbf{G}_t(\mathbf{x}_t^t)$ is a state-dependent, $n \times m$ matrix, where $\left\langle \mathbf{G}_t(\mathbf{x}_t^t)\mathbf{G}_t(\mathbf{x}_t^t)^T \right\rangle = \mathbf{Q}_t$ and δ_t is a m-dimensional process we'd like to approximate as white noise with zero mean (see Penland 2003

for caveats). $\mathbf{Q}_t$ is assumed known. In reality, the noise process may neither be white nor have zero mean, and the structure of $\mathbf{Q}_t$ may be difficult to estimate. For now, these complications are ignored. Substituting (6) into (5) produces a stochastic-dynamic model for the evolution of a model state:

$$\mathbf{x}^t_{t+1} = \mathcal{M}(\mathbf{x}^t_t) + \mathbf{G}_t(\mathbf{x}^t_t)\ \delta_t \tag{7}$$

This equation implies that even if the initial state is known precisely (an impossibility), future model states cannot since unknown random model errors are continually added.

Conceptually, the evolution of the pdf can be modeled with the Fokker-Planck equation (e.g., Gardiner 1985):

$$\frac{\partial P(\mathbf{x}^t_t)}{\partial t} = -\nabla \cdot \left[\mathcal{M}(\mathbf{x}^t_t)\ P(\mathbf{x}^t_t)\right] + \sum_{i,j} \frac{\partial^2}{\partial x^t_{t(i)} \partial x^t_{t(j)}} \left(\frac{\mathbf{Q}_t}{2}\right)_{ij} P(\mathbf{x}^t_t) \tag{8}$$

If $\mathbf{Q}_t$ is zero, then only the first term remains, and the Fokker-Planck equation reduces to the Liouville equation (Ehrendorfer 1994ab), a continuity equation for the conservation of probability. Probability thus diffuses with time according to the chaotic dynamics of the forecast model. The second term includes the effects of model error, including the increased diffusion of probability due to model uncertainty as well as noise-induced drift (Sardeshmukh et al. 2001).

c. Limitations of Bayesian data assimilation

Unfortunately, neither the update nor the forecast steps in Bayesian data assimilation are practical for real-world NWP applications without some simplification. For the update step, one problem with modeling a complicated pdf in higher dimensions is the "curse of dimensionality" (e.g., Bellman 1961, Hastie et al. 2001). Were one to try estimate the probability density in a higher-dimensional space using a small ensemble, one would find that the model of probability was very poor unless simplifying assumptions about the form of the distribution were made. Even were this problem surmountable, the computational cost would be extravagant. In the prior example the probability density was evaluated on a 100×100 grid. Suppose a similarly complicated structure for the prior existed in 100 dimensions. Then if the joint probabilities were monitored on a similar grid for each dimension, this would involve evaluating and modifying 100^{100}

density estimates. Such computations are already prohibitive for a 100-dimensional model state; the problem becomes incomprehensible for model states of $O(10^7)$.

Similarly, the Fokker-Planck equation is conceptually appealing but practically unusable for large systems; The pdf cannot be forecast using (8) in high-dimensional systems due to computational constraints.

There is a handy statistical property that suggests a solution to the problem. This property (e.g. Casella and Berger 1990, p. 45) states that when given a random variable such as $\mathbf{x}_t^t$, a function of this such as $\mathcal{E}(\mathbf{x}_t^t)$ is also a random variable. This indicates that if we can generate a random sample from $P(\mathbf{x}_t^t \mid \psi_t)$ and if $\mathcal{E}$ can properly represent the deterministic and stochastic effects in (7), we can generate a random sample of $\mathbf{x}_{t+1}^t$. This process is commonly known as *ensemble forecasting*. The presumption is that if the initial uncertainty is randomly sampled, it is possible to generate realistic random samples of forecast uncertainty. In section 4, we will return to consider how to utilize the ensemble forecast data in the update process.

3. KALMAN FILTERS

An approximation to Bayesian state estimation is now considered under assumptions of linearity of error growth and normality of error distributions. This approximate method is known as the *Kalman filter* (Kalman 1960, Kalman and Bucy 1961, Jazwinski 1970, Gelb 1974, Ghil 1989, Cohn 1997, Talagrand 1997, Daley 1997). There are two components of the Kalman filter, an update step, where the state estimate and an estimate of the forecast uncertainty are adjusted to new observations, and a forecast step, where the updated state and the uncertainty estimate are propagated forward to the time when the next set of observations become available. The novel aspect of the Kalman filter compared to, say, operational methods such as three-dimensional variational assimilation (3D-Var; Lorenc 1986, Parrish and Derber 1992) is that the error statistics of the forecast are dynamically estimated, taking into account the tangent-linear growth of errors and the uncertainty in the model itself.

Kalman filters assume linearity and normality, for the assimilation problem becomes much more tractable when these assumptions are made. Non-normality of the prior such as the bimodality in Fig. 1a is typically assumed to be uncommon in atmospheric data assimilation. Linear and normal assumptions may be inappropriate for atmospheric data assimilations of moisture, cloud cover, and other aspects of the model state that may be very sensitive to motions at small scales, where the time scale of predictability is small and errors grow and saturate rapidly.

a. Updating in the discrete Kalman filter

Start by making the following assumptions:

$$P(\mathbf{x}_t^t \mid \psi_{t-1}) \sim N(\mathbf{x}_t^b, \mathbf{P}_t^b) \propto exp\left[-\frac{1}{2}\,(\mathbf{x}_t^t - \mathbf{x}_t^b)^T \mathbf{P}_t^{b\,-1}(\mathbf{x}_t^t - \mathbf{x}_t^b)\right]. \tag{9}$$

That is, the probability density of the prior is distributed as a multivariate normal distribution with known mean background $\mathbf{x}_t^b$ and background-error covariance matrix $\mathbf{P}_t^b$. Similarly, following the discussion in section 2a, assume that the observation likelihood is distributed as a multivariate normal distribution with mean $\mathbf{y}_t$ and covariance $\mathbf{R}$:

$$P(\mathbf{y}_t \mid \mathbf{x}_t^t) \sim N(\mathbf{y}_t, \mathbf{R}) \propto exp\left[-\frac{1}{2}\,\Big(\mathbf{H}(\mathbf{x}_t^t) - \mathbf{y}_t\Big)^T \mathbf{R}^{-1}\Big(\mathbf{H}(\mathbf{x}_t^t) - \mathbf{y}_t\Big)\right]. \tag{10}$$

Here, $\mathbf{H}$ is assumed to be a linear "forward" operator that converts the model state to the observation type and location. Applying (3),

$$P(\mathbf{x}_t^t \mid \psi_t) \propto exp\left[-\frac{1}{2}\,\Big(\mathbf{x}_t^t - \mathbf{x}_t^b\Big)^T \mathbf{P}_t^{b\,-1}\Big(\mathbf{x}_t^t - \mathbf{x}_t^b\Big) - \frac{1}{2}\,\Big(\mathbf{H}(\mathbf{x}_t^t) - \mathbf{y}_t\Big)^T \mathbf{R}^{-1}\Big(\mathbf{H}(\mathbf{x}_t^t) - \mathbf{y}_t\Big)\right]. \tag{11}$$

Maximizing (11) is equivalent to minimizing the negative natural log of (11), i.e., to minimizing the functional $J(\mathbf{x}_t^t)$ according to

$$J(\mathbf{x}_t^t) = \frac{1}{2}\left[\Big(\mathbf{x}_t^t - \mathbf{x}_t^b\Big)^T \mathbf{P}_t^{b\,-1}\Big(\mathbf{x}_t^t - \mathbf{x}_t^b\Big) + \Big(\mathbf{H}(\mathbf{x}_t^t) - \mathbf{y}_t\Big)^T \mathbf{R}^{-1}\Big(\mathbf{H}(\mathbf{x}_t^t) - \mathbf{y}_t\Big)\right]. \tag{12}$$

This functional is a common starting point in the derivation of many assimilation schemes, from the Kalman filter to 3D-Var. Let's choose the value that minimizes this functional, providing a maximum-likelihood estimate of the state which blends the new observations and the prior. This

estimate will be called the "analysis," or $\mathbf{x}_t^a$. $\mathbf{x}_t^a$ can be found by the differentiating the functional in (12) with respect to $\mathbf{x}_t^t$, setting the result equal to zero, and proceeding with some manipulation. The resulting "update" equations are

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}\left(\mathbf{y}_t - \mathbf{H}(\mathbf{x}_t^b)\right), \tag{13}$$

where

$$\mathbf{K} = \mathbf{P}_t^b\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1}. \tag{14}$$

The optimal analysis state $\mathbf{x}_t^a$ is estimated by correcting the background $\mathbf{x}_t^b$ toward the "observation increment" $\mathbf{y}_t - \mathbf{H}(\mathbf{x}_t^b)$, weighted by the Kalman-gain matrix $\mathbf{K}$. In $\mathbf{K}$, $\mathbf{P}_t^b\mathbf{H}^T$ represents the covariances between the background state and the background state converted to observation location and variable type; $\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T$ represents the background-error covariance expressed at the observation location expressed in the units of the observations. The effect of $\mathbf{K}$ is to apply the observation increments to correct the background at relevant surrounding grid points. On average, the corrections are typically larger for grid points near to the observation location than for grid points far from the observation location. Further, the larger $\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T$ is compared to $\mathbf{R}$, the more the analysis is drawn to the observations.

Background-error statistics are explicitly updated in the Kalman filter. Given $\mathbf{P}_t^b$, $\mathbf{R}$, and the observation locations (implied in $\mathbf{H}$), the analysis-error covariance $\mathbf{P}_t^a$ is predicted. The equation for the expected $\mathbf{P}_t^a$ is

$$\begin{aligned}\mathbf{P}_t^a &= (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{P}_t^b \\ &= \mathbf{P}_t^b - \mathbf{K}\mathbf{H}\mathbf{P}_t^b \\ &= \mathbf{P}_t^b - \mathbf{P}_t^b\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_t^b.\end{aligned} \tag{15}$$

The accuracy of the mean and covariance updates will depend on whether the underlying assumption of normality is met and whether the error statistics have been estimated accurately. Where do these statistics come from? Background-error covariances $\mathbf{P}_t^b$ are estimated by evolving analysis-error covariances forward to the time of the next assimilation cycle, discussed more in section 3b. For atmospheric data assimilation, $\mathbf{R}$ is usually derived from extensive calibration

and validation during field experiments. **R** is usually assumed to be a combination of random instrument error and the error of representativeness, which can be thought of as the error in the interpolation operator **H** (Lorenc 1986). Error statistics for **R** are commonly assumed to be independent of the flow (Daley 1991, 1993). That is, the same error statistics are used for different observations of the same type, regardless of location, time, and weather conditions, an assumption that may not always be justified.

b. Forecasting the state and error covariances in the discrete Kalman filter

The appeal of the Kalman filter relative to an analysis scheme like 3D-Var is that the influence of new observations (as expressed in the structure of the analysis increments $\mathbf{x}_t^a - \mathbf{x}_t^b$) can be quite complicated and flow- and time-dependent. These complicated structures result from the model of background-error covariances developed by the Kalman filter. These error covariances in turn reflect the structure of prior analysis-error covariances, the observation network, how the flow-dependent linear error dynamics grows or shrinks these errors with time, and the effects of model uncertainty.

Suppose the analysis $\mathbf{x}_t^a$ and its error covariance $\mathbf{P}_t^a$ are known. An estimate will be needed for the subsequent model state and background-error covariances at time $t + 1$. In the discrete Kalman filter, suppose that the true model state evolves according to the linear equation

$$\mathbf{x}_{t+1}^t = \mathbf{M}(\mathbf{x}_t^t) + \eta. \tag{16}$$

That is, the forecast evolution can be expressed as the sum of a linear operation on the current model state $\mathbf{M}(\mathbf{x}_t^t)$ plus an unknown error η, also known as the "system noise." **M** is an $n \times n$ matrix, often called the *transition matrix* between times t and $t + 1$. Since the discrete Kalman filter is described only to motivate use of the ensemble Kalman filter, η is assumed to have an expected value of zero $\langle \eta \rangle = 0$ and to be uncorrelated in time, with expected "model error" covariance $\mathbf{Q}$: $\langle \eta_p \eta_q^T \rangle = \mathbf{Q}\delta_{pq}$ where δ is the Kronecker delta, and $_p$ and $_q$ denote two assimilation times. Algorithmic details like the method of calculation of **M** will be skipped; these details are

not crucial to the understanding of ensemble assimilation methods; for more details, see, for example, Le Dimet and Talagrand (1986) and Lacarra and Talagrand (1988).

Given the linear model (16), the estimate of the evolution of the mean state (e.g., Talagrand 1997) is simply (16) without the noise term:

$$\mathbf{x}^{b}_{t+1} = \mathbf{M}(\mathbf{x}^{a}_{t}). \tag{17}$$

An estimate of the background-error covariances is also required at the next assimilation time. Assuming that the system noise is uncorrelated with the tangent-linear dynamics, $\left\langle \mathbf{M}(\mathbf{x}^{a}_{t} - \mathbf{x}^{t}_{t})\ \eta^{T} \right\rangle = 0$, using (16) and (17) we get

$$\begin{aligned}
\mathbf{P}^{b}_{t+1} &= \left\langle (\mathbf{x}^{b}_{t+1} - \mathbf{x}^{t}_{t+1})\ (\mathbf{x}^{b}_{t+1} - \mathbf{x}^{t}_{t+1})^{T} \right\rangle \\
&= \left\langle \left(\mathbf{M}(\mathbf{x}^{a}_{t}) - \mathbf{M}(\mathbf{x}^{t}_{t}) - \eta \right)\ \left(\mathbf{M}(\mathbf{x}^{a}_{t}) - \mathbf{M}(\mathbf{x}^{t}_{t}) - \eta \right)^{T} \right\rangle \\
&= \left\langle \left(\mathbf{M}(\mathbf{x}^{a}_{t} - \mathbf{x}^{t}_{t}) - \eta \right)\ \left(\mathbf{M}(\mathbf{x}^{a}_{t} - \mathbf{x}^{t}_{t}) - \eta \right)^{T} \right\rangle \\
&= \mathbf{M}\mathbf{P}^{a}_{t}\mathbf{M}^{T} + \mathbf{Q} \\
&= \mathbf{M}\ (\mathbf{M}\mathbf{P}^{a}_{t})^{T} + \mathbf{Q}.
\end{aligned} \tag{18}$$

Given an operator **M**, this is how analysis-error covariances are evolved in the discrete Kalman filter.

In practice, accurately determining even the time-averaged statistics of **Q** may be quite complicated (Cohn and Parrish 1991, Daley 1992, Dee 1995, Blanchet et al. 1997). For both the Kalman filter and ensemble-based methods, the accuracy of the assimilation is likely to strongly depend on the assumed model for **Q**. Methods for estimating **Q** will be discussed for ensemble-based methods in section 4d.

c. The extended Kalman filter

Before considering ensemble-based data assimilation methods, an extension to the discrete Kalman filter called the *extended Kalman filter* (EKF) is considered (Jazwinski 1970, Gelb 1974, Gauthier et al. 1992, Bouttier 1994). In the EKF, some of the assumptions of linearity are relaxed. First, suppose the assumption of linearity in (16) is a poor one; perhaps were one to use a

fully nonlinear model operator $\mathcal{M}$, i.e.,

$$\mathbf{x}_{t+1}^{t} = \mathcal{M}(\mathbf{x}_{t}^{t}) + \eta, \tag{19}$$

then η would be much smaller. Accordingly, in the extended Kalman filter, one assumes that the mean state will be evolved according to

$$\mathbf{x}_{t+1}^{b} = \mathcal{M}(\mathbf{x}_{t}^{a}) \tag{20}$$

instead of using (17). If differences $\mathbf{x}_t^a - \mathbf{x}_t^t$ are small, then the evolution of these difference should be approximately linear:

$$\mathcal{M}(\mathbf{x}_{t}^{a}) - \mathcal{M}(\mathbf{x}_{t}^{t}) \simeq \mathbf{M}(\mathbf{x}_{t}^{a} - \mathbf{x}_{t}^{t}), \tag{21}$$

where here $\mathbf{M}$ is the Jacobian matrix of $\mathcal{M}$: $\mathbf{M} = \frac{\partial \mathcal{M}}{\partial \mathbf{x}}$.

The restriction that the forward operator $\mathbf{H}$ in (10) be linear is also relaxed, permitting a possibly nonlinear $\mathcal{H}$. Again, we presume that differences like $\mathbf{x}_t^b - \mathbf{x}_t^t$ are small enough so that the innovation vector $\mathbf{y}_t - \mathcal{H}(\mathbf{x}_t^b) = \mathcal{H}(\mathbf{x}_t^t) - \mathcal{H}(\mathbf{x}_t^b) + \epsilon$ can be approximated with $\mathbf{H}(\mathbf{x}_t^t - \mathbf{x}_t^b) + \epsilon$, where now $\mathbf{H} = \frac{\partial \mathcal{H}}{\partial \mathbf{x}}$.

Given these relaxed assumptions, one can proceed to derive an alternate form of the Kalman filter update equations. We will be content here to simply note the changes. In addition to assuming that the mean state is evolved nonlinearly in (20), the update equations is changed. (13) is replaced by

$$\mathbf{x}_{t}^{a} = \mathbf{x}_{t}^{b} + \mathbf{K}\Big(\mathbf{y}_t - \mathcal{H}(\mathbf{x}_{t}^{b})\Big), \tag{22}$$

where

$$\mathbf{K} = \mathbf{P}_{t}^{b}\mathbf{H}^{T}(\mathbf{H}\mathbf{P}_{t}^{b}\mathbf{H}^{T} + \mathbf{R})^{-1}. \tag{23}$$

Covariance propagation is done as in (18).

d. Considerations in the use of Kalman filters

What approximations may limit the accuracy of the Kalman filter? First, as mentioned previously, error statistics must be carefully estimated; in particular, $\mathbf{Q}$ may be especially problem-

atic. Second, error covariances are assumed to be normally distributed, so if they depart substantially from normality, Kalman filters may not perform well. There is also the potential disadvantage of the assumption of linearity of error growth. While errors for large-scale variables may reasonably be assumed to grow linearly over a typical period between assimilation cycles of 3-6 h, some aspects, especially moist thermodynamic variables, may have errors which saturate on this time scale. Similarly, if observations are not regularly available, error covariances estimated with tangent-linear dynamics may grow rapidly without bound (Evensen 1992, Gauthier et al. 1993, Bouttier 1994).

Another potential disadvantage of the Kalman filters may be their computational expense. Though Kalman filters provide a dramatic reduction in the computational cost relative to full Bayesian data assimilation, for a highly dimensional state vector, the computational costs in weather prediction models may still be extravagant. Consider the last line in (18). For an n-dimensional model state vector, it will require $2n$ applications of $\mathbf{M}$ to forecast the error covariances. Some reductions of computational expense may be possible. For example, there have been suggestions that this computation may be more practical if the tangent-linear calculations are performed in a subspace of the leading singular vectors (Fisher 1998, Farrell and Ioannou 2001).

Much more can be said about the Kalman filter, such as its equivalence to 4D-Var under certain assumptions (Li and Navon 2001), the manner of computing $\mathbf{M}$, iterated extensions of the basic extended Kalman filter (Jazwinski 1970, Gelb 1974, Cohn 1997), and the properties of its estimators (which, in the case of the discrete filter, if assumptions hold, provide the Best Linear Unbiased Estimate, or BLUE; see Talagrand 1997). These Kalman filters, however, are mentioned because they provide a context for understanding ensemble-based assimilation methods.

4. ENSEMBLE-BASED DATA ASSIMILATION

The Kalman filter is computationally expensive, and the assumptions that errors evolve linearly and are normally distributed may hinder its applicability to high-dimensional geophysical systems. Ensemble-based assimilation algorithms may be able to provide a more accurate analysis in situations where nonlinearity is strong and statistics exhibit some non-normality. If these assimilation algorithms can work accurately with many fewer ensemble members than elements in the state vector, then they will be computationally much less expensive as well. Consequently, many researchers have proposed a variety of ensemble-based assimilation methods that more perform more robustly when nonlinearity and non-normality may occur. Despite the many differences between the various ensemble-based algorithms, all are comprised of a finite number (perhaps 10 to a few hundred) parallel data assimilation and short-range forecast cycles. Background-error covariances are modeled using the ensemble of forecasts, and an ensemble of analyses are produced, followed by an ensemble of short-term forecasts to the next time observations are available. Ensemble-based assimilation algorithms also have the desirable property that if error dynamics are indeed linear and the error statistics Gaussian, then as the ensemble size increases, the state and covariance estimate from ensemble algorithms converge to those obtained from the extended Kalman filter.

The concepts behind ensemble assimilation methods have been used in engineering and aerospace applications as far back as the 1960's (Potter 1964, Kaminski et al. 1971, Maybeck 1979). Leith (1983) outlined the basic idea for atmospheric data assimilation. The idea was more completely described and tested in an oceanographic application by Evensen (1994) and in atmospheric data assimilation by Houtekamer and Mitchell (1998). Evensen (2003) provides a general overview of these algorithms, and Lorenc (2003) hypothesizes on the relative benefits and drawbacks relative to 4D-Var. Here, two classes of ensemble-based filters will be described, stochastic and deterministic. Both classes propagate the ensemble of analyses with nonlinear forecast models; the difference is whether or not random noise is applied during the update step.

a. Stochastic update algorithms

The most well-known ensemble-based data assimilation algorithm is the *ensemble Kalman filter*, or "EnKF" (Evensen 1994, Evensen and van Leeuwen 1996, Houtekamer and Mitchell 1998, 1999, 2001, Burgers et al. 1998, Hamill and Snyder 2000, 2002, Mitchell and Houtekamer 2000, Hamill et al. 2001, Keppenne and Rienecker 2002, Mitchell et al. 2002). This algorithm updates each member to a different set of observations perturbed with random noise. Because randomness is introduced every assimilation cycle, the update is considered stochastic.

For notational simplicity, the t time subscript used in previous sections is dropped; it is assumed unless noted otherwise that we are interested in estimating the state pdf at time t. We start off by assuming that we have an ensemble of forecasts that randomly sample the model background errors at time t. Let's denote this ensemble as $\mathbf{X}^{\mathrm{b}}$, a matrix whose columns are comprised of ensemble member's state vectors:

$$\mathbf{X}^{\mathrm{b}} = (\mathbf{x}_1^{\mathrm{b}}, \ldots, \mathbf{x}_m^{\mathrm{b}}), \tag{24}$$

The subscript now denotes the ensemble member. The ensemble mean $\overline{\mathbf{x}}^{\mathrm{b}}$ is defined as

$$\overline{\mathbf{x}}^{\mathrm{b}} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{x}_i^{\mathrm{b}}. \tag{25}$$

The perturbation from the mean for the ith member is $\mathbf{x}_i'^{\mathrm{b}} = \mathbf{x}_i^{\mathrm{b}} - \overline{\mathbf{x}}^{\mathrm{b}}$. Define $\mathbf{X}'^{\mathrm{b}}$ as a matrix formed from an ensemble of perturbations

$$\mathbf{X}'^{\mathrm{b}} = (\mathbf{x}_1'^{\mathrm{b}}, \ldots, \mathbf{x}_m'^{\mathrm{b}}) \tag{26}$$

and let $\hat{\mathbf{P}}^{\mathrm{b}}$ represent an estimate of $\mathbf{P}^{\mathrm{b}}$ from a finite ensemble

$$\hat{\mathbf{P}}^{\mathrm{b}} = \frac{1}{m-1}\mathbf{X}'^{\mathrm{b}}\mathbf{X}'^{\mathrm{b}^{\mathrm{T}}}. \tag{27}$$

The EnKF performs an ensemble of parallel data assimilation cycles, $i = 1, \ldots, m$, with each member updated to a somewhat different realization of the observations:

$$\mathbf{x}_i^{\mathrm{a}} = \mathbf{x}_i^{\mathrm{b}} + \hat{\mathbf{K}}\Big(\mathbf{y}_i - \mathcal{H}(\mathbf{x}_i^{\mathrm{b}})\Big). \tag{28}$$

$\mathcal{H}$ is the observation operator, which is permitted here to be a nonlinear operator. In (28), the $\mathbf{y}_i = \mathbf{y} + \mathbf{y}'_i$ are "perturbed observations," defined such that $\mathbf{y}'_i \sim N(0, \mathbf{R})$, and

$$\frac{1}{m}\sum_{i=1}^{m}\mathbf{y}'_i = 0. \tag{29}$$

The m sets of perturbed observations are thus created to update the m different background fields. Here, in (28), $\hat{\mathbf{K}} = \hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}}(\mathcal{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}} + \mathbf{R})^{-1}$, similar to the extended Kalman filter gain in (23) but permitting a possibly nonlinear $\mathcal{H}$ and estimating the background-error covariance from the ensemble. This second difference may provide a powerful advantage over the extended Kalman filter. Envision a situation where errors grow rapidly but saturate at low amplitude; the linear assumption of error growth in the extended Kalman filter will result in an overestimate of background error variance, but the differences among ensemble members will not grow without bound and thus should provide a more accurate model of the actual background-error statistics.

Why does the EnKF assimilate perturbed observations in (28) rather than unperturbed observations? Let $\mathbf{X}'^{\mathrm{a}}$ be a matrix of analysis ensemble member deviations from the analysis mean state, as (26) defined background deviations, and let $\hat{\mathbf{P}}^{\mathrm{a}}$ be formed from the ensemble of analyses using (28). Then as the ensemble size approaches infinity and if the dynamics are linear, $\hat{\mathbf{P}}^{\mathrm{a}} = \frac{1}{m-1}\mathbf{X}'^{\mathrm{a}}\mathbf{X}'^{\mathrm{a}\mathrm{T}} \rightarrow \mathbf{P}^{\mathrm{a}}$, where $\mathbf{P}^{\mathrm{a}}$ is the extended Kalman filter analysis-error covariance from (15) (Burgers et al. 1998). If unperturbed observations are assimilated in (28) without other modifications to the algorithm, the analysis-error covariance will be underestimated, and observations will not be properly weighted in subsequent assimilation cycles. Pham (2001) proposes an alternative to perturbing the observations, adding noise to background forecasts in a manner that also ensures analysis-error covariances are equal to those produced by the extended Kalman filter.

For a complex numerical weather prediction model with a high-dimensional state vector, explicitly forming $\hat{\mathbf{P}}^{\mathrm{b}}$ as in (27) would be computationally prohibitive; for example, in a model with 10^7 elements in its state, storing and readily accessing the 10^{14} elements of $\hat{\mathbf{P}}^{\mathrm{b}}$ is not possible. However, in ensemble-based methods, $\hat{\mathbf{K}}$ can be formed without ever explicitly computing

the full $\hat{\mathbf{P}}^{\mathrm{b}}$. Instead, the components of $\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}}$ and $\mathcal{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}}$ of $\hat{\mathbf{K}}$ are computed separately. Define

$$\overline{\mathcal{H}(\mathbf{x}^{\mathrm{b}})} = \frac{1}{m}\sum_{i=1}^{n}\mathcal{H}(\mathbf{x}_{\mathrm{i}}^{\mathrm{b}}),$$

which represents the mean of the estimate of the observation interpolated from the background forecasts. Then

$$\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}} = \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathbf{x}_i^{\mathrm{b}} - \overline{\mathbf{x}^{\mathrm{b}}}\right)\left(\mathcal{H}(\mathbf{x}_i^{\mathrm{b}}) - \overline{\mathcal{H}(\mathbf{x}^{\mathrm{b}})}\right)^{\mathrm{T}}, \tag{30}$$

and

$$\mathcal{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}} = \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathcal{H}(\mathbf{x}_i^{\mathrm{b}}) - \overline{\mathcal{H}(\mathbf{x}^{\mathrm{b}})}\right)\left(\mathcal{H}(\mathbf{x}_i^{\mathrm{b}}) - \overline{\mathcal{H}(\mathbf{x}^{\mathrm{b}})}\right)^{\mathrm{T}}. \tag{31}$$

Of course, if the number of observations is as large as the elements in the model state, $\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}}$ and $\mathcal{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}}$ will be as large as $\hat{\mathbf{P}}^{\mathrm{b}}$, negating this advantage. Section 6 will describe how the array size can be limited through the serial processing of independent batches of observations.

b. Deterministic update algorithms.

There are ways to correct the background ensemble to new observations so that $\hat{\mathbf{P}}^{\mathrm{a}} \rightarrow \mathbf{P}^{\mathrm{a}}$ without adding random noise. As discussed in Whitaker and Hamill (2002), adding noise to the observations may have several deleterious consequences, such as introducing spurious observation-background error correlations. Algorithms that do not add stochastic noise are called *deterministic* algorithms, so named because if the background ensemble and the associated error statistics are known, the ensemble of analysis states will be completely known as well. Such algorithms (e.g., Bishop et al. 2001, Anderson 2001, Whitaker and Hamill 2002) all attempt to adjust the existing ensemble in a way that provides the correct posterior covariances without adding noise to the observations or the background forecasts. Tippett et al. (2003) describes the similarities and differences between these algorithms. In each algorithm, the background-error covariances are never explicitly formed, with manipulations being performed on the matrix square root (i.e., eq (26), the matrix of ensemble member deviations from the mean). As pointed out in Tippett et al., since $\hat{\mathbf{P}}^{\mathrm{b}} = \frac{1}{m-1}\mathbf{X}'^{\mathrm{b}}\mathbf{X}'^{\mathrm{b}\mathrm{T}}$, given a matrix $\mathbf{U}$ representing an $n \times n$ orthogonal transformation $\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$, then $\hat{\mathbf{P}}^{\mathrm{b}}$ can also be represented as $\hat{\mathbf{P}}^{\mathrm{b}} = \frac{1}{m-1}(\mathbf{X}'^{\mathrm{b}}\mathbf{U})(\mathbf{X}'^{\mathrm{b}}\mathbf{U})^{\mathrm{T}}$. Hence, many square-root filters can be formulated that produce the same posterior covariance model.

A particularly simple implementation of a deterministic ensemble assimilation algorithm is the "ensemble square-root filter," or "EnSRF," described by Whitaker and Hamill (2002). Like the EnKF, the EnSRF conducts a set of parallel data assimilation cycles. It is convenient in the EnSRF to update the equations for the ensemble mean (denoted by an overbar) and the deviation of the ith member from the mean separately:

$$\overline{\mathbf{x}}^{\mathrm{a}} = \overline{\mathbf{x}}^{\mathrm{b}} + \hat{\mathbf{K}} \left(\mathbf{y} - \mathcal{H}(\overline{\mathbf{x}}^{\mathrm{b}}) \right), \tag{32}$$

$$\mathbf{x}_i'^{\mathrm{a}} = \mathbf{x}_i'^{\mathrm{b}} - \widetilde{\mathbf{K}} \mathcal{H}(\mathbf{x}_i'^{\mathrm{b}}). \tag{33}$$

Here, $\hat{\mathbf{K}}$ is the traditional Kalman gain as in Eq. (25), and $\widetilde{\mathbf{K}}$ is the "reduced" gain used to update deviations from the ensemble mean.

If observations have independent errors, they can be assimilated simultaneously or serially (sequentially), producing the same result (Gelb 1974, Bishop et al. 2001, and section 6a, below). The analysis ensemble after the assimilation of the first observation is used as the background ensemble for the assimilation of the second, and so on. When sequentially processing independent observations, $\hat{\mathbf{K}}$, $\widetilde{\mathbf{K}}$, $\mathcal{H}\hat{\mathbf{P}}^{\mathrm{b}}$ and $\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}}$ are all n-dimensional vectors, and $\mathcal{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}}$ and $\mathbf{R}$ are scalars. Thus, as first noted by Potter (1964), when observations are processed one at a time,

$$\widetilde{\mathbf{K}} = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathcal{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}} + \mathbf{R}}} \right)^{-1} \hat{\mathbf{K}}. \tag{34}$$

The quantity multiplying $\hat{\mathbf{K}}$ in Eq. (34) thus becomes a scalar between 0 and 1. This means that, in order to obtain the correct analysis-error covariance with unperturbed observations, a modified Kalman gain that is reduced in magnitude relative to the traditional Kalman gain is used to update deviations from the ensemble mean. Consequently, deviations from the mean are reduced less in the analysis using $\widetilde{\mathbf{K}}$ than they would be using $\hat{\mathbf{K}}$. In the EnKF, the excess variance reduction caused by using $\hat{\mathbf{K}}$ to update deviations from the mean is compensated for by the introduction of noise to the observations.

In the EnSRF, the mean and departures from the mean are updated independently according to Eqs. (32) and (33). If observations are processed one at a time, the EnSRF requires about

the same computation as the traditional EnKF with perturbed observations, but for moderately sized ensembles and processes that are generally linear and Gaussian, the EnSRF produces analyses with significantly less error (Whitaker and Hamill 2002). Conversely, Lawson and Hansen (2003) suggest that if nonlinearity and non-normality are typical and ensemble size is large, the EnKF will perform better.

Another deterministic update algorithm is the Ensemble Transform Kalman filter (ETKF) of Bishop et al. (2001). The ETKF finds the transformation matrix $\mathbf{T}$ such that $\hat{\mathbf{P}}^{\mathrm{a}} = \frac{1}{m-1}(\mathbf{X}'^{\mathrm{b}}\mathbf{T})\,(\mathbf{X}'^{\mathrm{b}}\mathbf{T})^{\mathrm{T}} \rightarrow \mathbf{P}^{\mathrm{a}}$. (see Bishop et al. for details on the computation of $\mathbf{T}$). Compared with the EnSRF, an advantage of the ETKF is its computational speed; a disadvantage is that the ETKF cannot apply covariance localizations (section 5), which may make the analyses very inaccurate unless large ensembles are used. The ETKF has been successfully demonstrated for generating perturbed initial conditions for ensemble forecasts about a mean state updated using 3D-Var (Wang and Bishop 2003), and hybrid ETKF-variational schemes are being explored (Etherton and Bishop 2003).

c. A simple demonstration of stochastic and deterministic update steps

Consider again the Bayesian data assimilation problem illustrated in Fig. 1. There, a bimodal 2-D probability distribution was updated to an observation of one component. Let's explore the characteristics of the EnKF and EnSRF update applied to this problem.

A realistic 100-member random sample was first generated from the bimodal pdf in Fig. 2a. These samples are denoted by the black dots in Fig. 2a. Let's keep track of the assimilation for one particular member, denoted by the larger black dot.

The EnKF and EnSRF adjust the background to the observations with weighting factors that assume the distributions are normal. Estimated from this random sample, the background-error covariance is

$$\hat{\mathbf{P}}^{\mathrm{b}} = \begin{pmatrix} \sigma^2(x^{\mathrm{b}}_{(1)}) & Cov\,(x^{\mathrm{b}}_{(1)}, x^{\mathrm{b}}_{(2)}) \\ Cov\,(x^{\mathrm{b}}_{(1)}, x^{\mathrm{b}}_{(2)}) & \sigma^2(x^{\mathrm{b}}_{(2)}) \end{pmatrix} \simeq \begin{pmatrix} 150.73 & 109.70 \\ 109.70 & 203.64 \end{pmatrix}.$$

The shape of this distribution is illustrated by the black contours in Fig. 2a. Here, the observation measures the same aspect as the first component of our state variable: $\mathcal{H} = [1, 0]$. As in Fig. 1, assume $\mathbf{R} = 100$, so $\mathcal{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}} + \mathbf{R} \simeq 150.73 + 100.00 = 250.73$. $\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}} \simeq [150.73, 109.70]^{\mathrm{T}}$, and hence $\hat{\mathbf{K}} = \mathbf{P}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}}(\mathcal{H}\mathbf{P}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}} + \mathbf{R})^{-1} \simeq [0.60, 0.44]^{\mathrm{T}}$.

For the EnKF, perturbed observations were then generated, denoted by the short vertical lines along the abscissa in Fig. 2a. Eq. (28) was then applied, updating background samples to their associated perturbed observations, generating analysis samples. For example, the heavy black dot in Fig. 2a was updated to the perturbed observation marked with the "*". The resulting analysis sample is the large black dot in Fig. 2b. For the noted sample, the first component of the background state was much less than the mean, and the perturbed observation was greater than the mean background state. The resulting analysis nudged the posterior state toward the mean in both components. Other dots in Fig. 2b denote other updated EnKF member states.

In the EnSRF, the ensemble background mean state $\sim [47.93, 50.07]^{\mathrm{T}}$ was updated to the mean observed value 58.0 using $\hat{\mathbf{K}}$ computed above and eq. (32), resulting in a mean analyzed state of $\sim [53.55, 54.16]$. As with the EnKF, given the positive observation increment and the positive correlation of the background-error covariances between the two components, both components of the mean state were adjusted upward. EnSRF perturbations from the mean were updated using eq (33) and the reduced gain, here $\widetilde{\mathbf{K}} \simeq 0.613\,\hat{\mathbf{K}}$.

Compare the EnKF and EnSRF random samples of the posterior from Figs. 2b-c and their fitted distribution (black contours) with the correct Bayesian posterior (red contours). The samples from both distributions do not appear to randomly sample the correct posterior. The EnKF and EnSRF posterior distributions are shifted slightly toward lower values in both components. The EnSRF posterior samples preserve the original shape from the prior, though their values are shifted in mean and compressed together. In comparison, the EnKF samples are randomized somewhat through the assimilation of the perturbed observations, and in this case, its distribution is somewhat more diffuse than that of the EnSRF. The EnKF samples appear to overlap more with the correct distribution than the samples from the EnSRF.

Why can't ensemble-based methods correctly adjust the prior ensemble to the new observations so that the samples reflect a random draw from the Bayesian posterior? The reason is that ensemble-based methods implicitly assume a second-moment closure; that is, the distributions are assumed to be fully described by means and covariances. The example shown above demonstrates that some inaccuracies can be expected in these analyses if indeed there are higher-moment details in these distributions. Lawson and Hansen (2003) show that the EnKF may be more robust than the EnSRF in such situations, while Whitaker and Hamill (2002) indicate that the EnSRF may perform better in situations where normality is an appropriate assumption, as the random noise introduced by perturbing the observations in the EnKF can have deleterious effects. Hopefully, highly non-normal distributions are not frequently encountered, as radically more expensive techniques than those discussed here may then be required (e.g., Gordon et al. 1993).

d. Ensemble propagation of the pdf and model-error parameterization

Ensemble-based assimilation methods leverage a potential increase in accuracy that may result from estimating covariances from an ensemble propagated with the fully nonlinear forecast model. If forecast-error dynamics are in fact quite nonlinear and saturate quickly, then the assumption of linearity in the discrete and extended Kalman filters was inappropriate, and some accuracy may be gained relative to the Kalman filters by estimating covariances from a sample of fully nonlinear model forecasts.

However, in real-world applications, background-error covariances cannot simply be estimated at the next assimilation cycle by conducting an ensemble of deterministic forecasts forward from the current cycle's analyses. Because of model deficiencies, even if the true state of the atmosphere is perfectly known, the resulting forecast will be imperfect: $\mathbf{x}^t_{(t+1)} = \mathcal{M}(\mathbf{x}^t_{(t)}) + \eta$, where here we denote the time index in parentheses and $\mathcal{M}$ is again the nonlinear forecast operator. Let's first assume that our forecast model is unbiased $\left\langle \eta \right\rangle = 0$, again with model-error covariance $\left\langle \eta\eta^{\mathrm{T}} \right\rangle = \mathbf{Q}$. In practice, The assumption of no bias is probably not justified, and if the bias can be determined, the forecasts ought to be corrected for this bias (Dee and Todling 2000) or more ideally, the forecast model ought to be improved. In any case, consider the error covari-

ance at the next assimilation time. Assume again that forecast error due to initial-condition uncertainty and model error are uncorrelated $\left\langle \left(\mathcal{M}(\overline{\mathbf{x}}^{\mathrm{a}}_{(\mathrm{t})}) - \mathcal{M}(\mathbf{x}^{\mathrm{t}}_{(\mathrm{t})}) \right) \eta^{\mathrm{T}} \right\rangle = 0$, and assume linearity of the error growth $\mathcal{M}(\overline{\mathbf{x}}^{\mathrm{a}}_{(\mathrm{t})}) - \mathcal{M}(\mathbf{x}^{\mathrm{t}}_{(\mathrm{t})}) \simeq \mathbf{M}\left(\overline{\mathbf{x}}^{\mathrm{a}}_{(\mathrm{t})} - \mathbf{x}^{\mathrm{t}}_{(\mathrm{t})}\right)$. Then the true background-error covariance at the next assimilation time is

$$\begin{aligned}
\left\langle \left(\overline{\mathbf{x}}^{\mathrm{b}}_{(\mathrm{t}+1)} - \mathbf{x}^{\mathrm{t}}_{(\mathrm{t}+1)}\right) \left(\overline{\mathbf{x}}^{\mathrm{b}}_{(\mathrm{t}+1)} - \mathbf{x}^{\mathrm{t}}_{(\mathrm{t}+1)}\right)^{\mathrm{T}} \right\rangle &= \left\langle \left(\mathcal{M}(\overline{\mathbf{x}}^{\mathrm{a}}_{(\mathrm{t})}) - \mathcal{M}(\mathbf{x}^{\mathrm{t}}_{(\mathrm{t})}) - \eta \right) \left(\mathcal{M}(\overline{\mathbf{x}}^{\mathrm{a}}_{(\mathrm{t})}) - \mathcal{M}(\mathbf{x}^{\mathrm{t}}_{(\mathrm{t})}) - \eta \right)^{\mathrm{T}} \right\rangle \\
&\simeq \left\langle \mathbf{M}\left(\overline{\mathbf{x}}^{\mathrm{a}}_{(\mathrm{t})} - \mathbf{x}^{\mathrm{t}}_{(\mathrm{t})}\right)\left(\overline{\mathbf{x}}^{\mathrm{a}}_{(\mathrm{t})} - \mathbf{x}^{\mathrm{t}}_{(\mathrm{t})}\right)^{\mathrm{T}} \mathbf{M}^{\mathrm{T}} \right\rangle + \left\langle \eta\eta^{\mathrm{T}} \right\rangle \\
&= \mathbf{M}\mathbf{P}^{\mathrm{a}}_{(\mathrm{t})}\mathbf{M}^{\mathrm{T}} + \mathbf{Q}
\end{aligned} \tag{35}$$

where $\mathbf{M}$ is again the Jacobian of the nonlinear operator. Consider what happens when covariances are estimated directly from an ensemble of forecasts propagated forward from an ensemble of $i = 1, \ldots, m$ analyses using the fully nonlinear forecast model

$$\mathbf{x}^{\mathrm{b}}_{\mathrm{i}(t+1)} = \mathcal{M}(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)}), \tag{36}$$

Calculating the expected covariance, we get

$$\begin{aligned}
\left\langle \left(\mathbf{x}^{\mathrm{b}}_{\mathrm{i}(t+1)} - \overline{\mathbf{x}}^{\mathrm{b}}_{(t+1)}\right) \left(\mathbf{x}^{\mathrm{b}}_{\mathrm{i}(t+1)} - \overline{\mathbf{x}}^{\mathrm{b}}_{(t+1)}\right)^{\mathrm{T}} \right\rangle &= \left\langle \left(\mathcal{M}(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)}) - \mathcal{M}(\overline{\mathbf{x}}^{\mathrm{a}}_{(t)})\right)\left(\mathcal{M}(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)}) - \mathcal{M}(\overline{\mathbf{x}}^{\mathrm{a}}_{(t)})\right)^{\mathrm{T}} \right\rangle \\
&\simeq \left\langle \mathbf{M}\left(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)} - \overline{\mathbf{x}}^{\mathrm{a}}_{(t)}\right)\left(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)} - \overline{\mathbf{x}}^{\mathrm{a}}_{(t)}\right)^{\mathrm{T}} \mathbf{M}^{\mathrm{T}} \right\rangle \\
&\simeq \mathbf{M}\hat{\mathbf{P}}^{\mathrm{a}}_{(\mathrm{t})}\mathbf{M}^{\mathrm{T}}.
\end{aligned} \tag{37}$$

Comparing (35) and (37), it is apparent that an ensemble of analyses that are simply propagated forward with the nonlinear forecast model will have too small an expected amount of spread, missing the extra model-error covariance $\mathbf{Q}$. Let us define some hypothetical set of background forecasts at time $t + 1$ that *do* have the correct covariance, i.e., define $\check{\mathbf{x}}^{\mathrm{b}}_{\mathrm{i}(t+1)}$ such that $\left\langle \left(\check{\mathbf{x}}^{\mathrm{b}}_{\mathrm{i}(t+1)} - \overline{\check{\mathbf{x}}^{\mathrm{b}}}_{(t+1)}\right) \left(\check{\mathbf{x}}^{\mathrm{b}}_{\mathrm{i}(t+1)} - \overline{\check{\mathbf{x}}^{\mathrm{b}}}_{(t+1)}\right)^{\mathrm{T}} \right\rangle = \mathbf{M}\hat{\mathbf{P}}^{\mathrm{a}}_{(\mathrm{t})}\mathbf{M}^{\mathrm{T}} + \mathbf{Q}$. Such an ensemble is possible if we add noise to our existing ensemble:

$$\check{\mathbf{x}}^{\mathrm{b}}_{\mathrm{i}(t+1)} = \mathbf{x}^{\mathrm{b}}_{\mathrm{i}(t+1)} + \xi_{\mathrm{i}}, \tag{38}$$

where $\left\langle \xi_{\mathrm{i}}\xi_{\mathrm{i}}^{\mathrm{T}} \right\rangle = \mathbf{Q}$, $\left\langle \xi_{\mathrm{i}} \right\rangle = 0$, and $\left\langle \mathbf{x}^{\mathrm{b}}_{\mathrm{i}(t+1)}\, \xi_{\mathrm{i}}^{\mathrm{T}} \right\rangle = 0$.

Several methods have been proposeed for incorporating noise into the ensemble of forecasts so that they account for model error. First, the forecast model could be stochastic-dynamic in-

stead of deterministic, with terms to the prognostic equations to represent interactions with unresolved scales and/or mis-parameterized effects; in essence, $\mathcal{M}$ is changed so that the ensemble of forecasts integrates random noise in addition to the deterministic forecast dynamics, as in (7). Over an assimilation cycle, this additional variance added to the ensemble as a result of integrating noise should be designed to increase the covariance by the missing **Q**. Another possibility is that one may choose to run a forecast model without integrating noise but to add noise to each member at the data assimilation time so as to increase the ensemble variance appropriate to the missing **Q**. Third, it may be possible to use a multi-model ensemble to estimate covariances, or to achieve satisfactory results by inflating the deviations of ensemble members about their mean.

Little work has yet been done on the first of these three approaches. Buizza et al. (1999) demonstrated a simple technique for integrating noise to account for deterministic sub-gridscale parameterizations. Under their methodology, the parameterized terms in the prognostic equations were multiplied by a random number. Penland (2003) outlines a more general approach for integrating system noise in numerical models. To date, however, a comprehensive noise integration scheme has not yet been demonstrated in an operational weather prediction model. Palmer (2001) discusses the potential appeal of such an approach.

The second general approach is to augment the ensemble-estimated model of covariances during the update step with noise representing the missing model error covariances. Mitchell and Houtekamer (2000) describe one such approach.

A third approach, use of multiple forecast models for generating the ensemble of background forecasts (e.g., Houtekamer et al. 1996b, Harrison et al. 1999, Evans et al. 2000, Ziehmann 2000, Richardson 2000, Hou et al. 2001), is appealing for its simplicity. A wider range of forecasts is typically generated when different weather forecast models are used to forecast the evolution of different ensemble members. Unfortunately, it is not clear whether or not the differences between members are actually representative of model errors; initial experimentation has shown that the multi-model ensembles tend to produce unrealistic estimates of error covariances. Forecast errors at larger scales ought to be mostly in balance, but when estimated from multi-

model ensembles, preliminary results suggest that the errors can be greatly out of balance, with detrimental effects on the subsequent assimilation (personal communication, M. Buehner). See also Hansen (2002) for a discussion of discussion of the use of multi-model ensembles in data assimilation in a simple model.

A last approach, discussed in Anderson and Anderson (1999) and demonstrated in Whitaker and Hamill (2002) and Whitaker et al. (2003), is to enlarge background error estimates by inflating forecast ensemble member's deviation about their mean by an amount r (slightly greater than 1.0) before the first observation is assimilated:

$$\mathbf{x}_i^b \leftarrow r\left(\mathbf{x}_i^b - \overline{\mathbf{x}}^b\right) + \overline{\mathbf{x}}^b. \tag{39}$$

Here, the operation $\leftarrow$ denotes a replacement of the previous value of $\mathbf{x}_i^b$. This technique is called "covariance inflation." Application of a moderate inflation factor has been found to improve the accuracy of assimilations. Note that inflation effectively increases the spread of the ensemble, but it does not change the subspace spanned by the ensemble. Hence, if model error projects into a substantially different subspace, this parameterization may not be effective.

5. COVARIANCE LOCALIZATION

In ensemble assimilation methods, the accuracy of error statistics is especially important. Unlike 3D-Var, the effects of a mis-specification of error statistics can affect the analysis-error covariance, which is then propagated forward in time. Hence, if the analysis errors are underestimated in one cycle, the forecast errors may be underestimated in the following cycle, underweighting the new observations. The process can feed back on itself, the ensemble assimilation method progressively ignoring observational data more and more in successive cycles, leading eventually to a useless ensemble. This is known as *filter divergence* (e.g., Houtekamer and Mitchell 1998, van Leeuwen 1999, Hamill et al. 2001). For the ensemble-based methods, filter divergence can be induced by many causes. One of the most crucial is to model background-error covariance realistically. As discussed in the previous section, an adequate parameterization of model error is likely to be very important. However, filter divergence can occur even in sim-

ulations where the forecast model is perfect, for background-error covariances are typically estimated imperfectly from small ensembles. While more ensemble members would be desirable to reduce the sampling error in estimating background-error covariances, more members means more computational expense.

One common algorithmic modification to improve error covariance estimates from ensembles is *covariance localization*. The covariance estimate from the ensemble is multiplied point by point with a correlation function that is 1.0 at the observation location and decreases monotonically with increasing distance. Why do this? Consider the example illustrated by Fig. 3. Here there is a 100-dimensional state vector, perhaps representing, say, surface temperature around a latitude circle. Suppose the true covariance is known, with a variance at each model grid point of 1.0 and a standard deviation of 7.5 grid points. In this case, errors $\sim$ 20 grid points apart are effectively uncorrelated (solid lines in Fig. 3a). When the covariance structure is estimated from a 50-member ensemble constructed to randomly sample the true covariance matrix (Houtekamer 1993, eq. 13), variances are mis-estimated somewhat, and there are spurious covariances between distant grid points. For example, grid point 40 and grid point 90 co-vary, as indicated by the dashed blue line. Hence, without some modification, an observation at grid point 40 will improperly change the analysis and reduce the posterior variance at grid point 90. After the application of a localization function (Fig. 3b) with a correlation length of 15 grid points, the covariance estimate much more closely resembles the true covariances.

Mathematically, to apply covariance localization, the Kalman gain $\hat{\mathbf{K}} = \hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}}(\mathcal{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}} + \mathbf{R})^{-1}$ is replaced by a modified gain

$$\hat{\mathbf{K}} = \left(\rho_S \circ \hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}}\right)\left(\mathcal{H}(\rho_S \circ \hat{\mathbf{P}}^{\mathrm{b}})\mathcal{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}, \tag{40}$$

where the operation $\rho_S \circ$ in (40) denotes a Schur product (an element-by-element multiplication) of a correlation matrix $\mathbf{S}$ with local support with the covariance model generated by the ensemble. One such correlation matrix can be constructed using an approximately Gaussian-shaped function described in Gaspari and Cohn (1999). Define a correlation length scale l_c, and let $F_c = \sqrt{\frac{10}{3}}\, l_c$. Define $\|D_{ij}\|$ to be the Euclidean distance between grid points i and j. Then a

correlation matrix **S** is defined according to $S(i,j) = \Omega\left(F_c, \|D_{ij}\|\right)$. Let $a = F_c$ and $b = \|D_{ij}\|$. Then

$$\Omega(a,b) = \begin{cases} 1 - \frac{1}{4}\left(\frac{b}{a}\right)^5 + \frac{1}{2}\left(\frac{b}{a}\right)^4 + \frac{5}{8}\left(\frac{b}{a}\right)^3 - \frac{5}{3}\left(\frac{b}{a}\right)^2, & 0 \le b \le a; \\ \frac{1}{12}\left(\frac{b}{a}\right)^5 - \frac{1}{2}\left(\frac{b}{a}\right)^4 + \frac{5}{8}\left(\frac{b}{a}\right)^3 + \frac{5}{3}\left(\frac{b}{a}\right)^2 - 5\left(\frac{b}{a}\right) + 4 - \frac{2}{3}\left(\frac{b}{a}\right)^{-1}, & a < b \le 2a; \\ 0. & b > 2a. \end{cases} \quad (41)$$

The Schur product of matrices **A** and **B** is a matrix **C** of the same dimension, where $C_{ij} = A_{ij}B_{ij}$. When covariance localization is applied to smaller ensembles, it can actually result in more accurate analyses than would be obtained from larger ensembles without localization (e.g., Houtekamer and Mitchell 2001). Mathematically, localization increases the effective rank of the background-error covariances (Hamill et al. 2001). In the extreme, if the correlation matrix **S** were the identity matrix, the covariance model would consist of grid points with variances and zero covariance and the rank of the covariance matrix after localization would increase from $m - 1$ to n, the dimension of the state vector. In practice, such an extreme localization would harm the quality of the analysis, destroying the mass-wind balance (Mitchell and Houtekamer 2002, Lorenc 2003) and prohibiting the observation from changing the analysis at nearby grid points. Hence, broader localizations are typically used. Generally, the larger the ensemble, the broader the optimum correlation length scale of the localization function (Houtekamer and Mitchell 2001, Hamill et al (2001)).

In applying the covariance localization, the covariances between distant grid points are effectively de-coupled. Should they be? Consider a simple two-member ensemble; dynamically, there is no reason to expect that the growth of differences over Japan is dynamically interconnected to the growth of differences over Africa, and neither interconnected with the growth of differences over South America. This two-member ensemble "sees" more than two growing regions but assumes they are all coupled. Covariance localization is thus a heuristic attempt to modify the model of background-error covariances so that a limited-size ensemble will not represent distant, distinct features as dynamically interrelated when in fact they generally are not. If indeed distant regions are in fact dynamically coupled, the localization will cause the loss of this information. The effect on the data assimilation will be that observations will not be able to

change the analysis and reduce the analysis-error variance in distant regions; local observations will have to be relied upon instead. This is judged to be less detrimental than the opposite, to let observations affect distant regions when this is inappropriate. It is especially important to guard against underestimating background-error covariances with ensembles, since this can induce filter divergence. Hence, the conservative use of the observations implied by localization is often helpful.

Covariance localization will be illustrated in a real-data example in section 7. For more background and experimentation, see Houtekamer and Mitchell (2001) and Hamill et al. (2001).

6. MAKING ENSEMBLE ASSIMILATION METHODS COMPUTATIONALLY FEASIBLE

Though ensemble-based assimilation is computationally less expensive than the Kalman filter, it is still a much more expensive calculation than 3D-Var and it is probably roughly the same order of magnitude as 4D-Var. The computational expense of most ensemble-based methods will scale with the number of observations times the number of ensemble members times the dimension of the state vector. Practically, though, the relative expense may be more determined by factors such as the convergence rate of 4D-Var, the extent of parallelization of ensemble approaches, and the complexity of the operator $\mathcal{H}$.

How can the computational time of these ensemble assimilation methods be lessened? One shortcut was already mentioned, avoiding the formulation of $\hat{\mathbf{P}}^{\mathrm{b}}$ directly and calculating the gain components directly from the ensemble. (eqs. 30 and 31). We now consider three other methods for simplifying the calculations, serial (sequential) processing of observations and parallel-processing methods, and variational formulation.

a. Serial processing of observations

Gelb (1974) demonstrated in the Kalman filter that a batch of independent observations can be assimilated either simultaneously or sequentially (serially). Starting with a background forecast and background error-covariances, a single observation may be assimilated. The analysis

state and analysis-error covariance resulting from the assimilation of the first observation may be used as the background and background-error covariance in the assimilation of the second independent observation. Regardless of whether the observation are assimilated simultaneously or serially, the same expected analysis and analysis-error covariance will result. Consequently, the order in which observations are serially assimilated is unimportant.

Serial processing of observations may be desirable in ensemble-based methods, for both the coding and the computations are rendered somewhat less demanding. Most of the ensemble-based algorithms described in the literature serially assimilate observations. When observations are assimilated serially, for each observation that is assimilated, $\mathcal{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}}$ and $\mathbf{R}$ become scalars. Thus, the inverse $(\mathcal{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}} + \mathbf{R})^{-1}$ in the gain matrix is trivial to compute. Also, the application of the covariance localization in (40) is much more straightforward to apply.

The equivalence of serial and simultaneous processing is only true if observations have independent errors (Kaminsky et al. 1971). Practically, however, many observations may have vertically or horizontally correlated errors. Consider two alternatives to deal with this. First, if the size of a batch of observations with correlated errors is relatively small, these batches can be processed simultanteously without much more computational expense; the matrix inverse of $(\mathcal{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}} + \mathbf{R})^{-1}$ should not be prohibitively expensive. Another option is to transform the observations and the forward operator so that the observations are effectively independent. The method for doing this is derived in the Appendix.

b. Parallel processing

Many modern computers today are "massively parallel," consisting of individual central processing units (CPUs) with their own memory. Such computers often have relatively fast computational speed on each CPU, but the overall speed of a calculation can be slowed significantly when new data must be continually exchanged between CPUs. Hence, to speed up the processing speed of ensemble data assimilation on such computers, the parallelization ought to be designed so a minimum of data need be exchanged among the CPUs.

There are two general components to the ensemble data assimilation; first, for an m-member ensemble-based method, m parallel data assimilation cycles will need to be computed, and second, from the m resulting analyses, m forecasts forward to the next assimilation time. The forecast component is easily parallelizable. Each member of the ensemble can be forecast in parallel on a separate CPU(s), since no information need be swapped between CPUs during the forecast step. However, parallelizing the data assimilation component is considerably more difficult. The update of each member background forecast to the new observations requires information from all the other ensemble members in formulating the gain matrix $\hat{\mathbf{K}}$. It would be highly inefficient to simply parallelize so that different members were updated on different processors, since each processor would be duplicating the same computationally expensive gain calculation. How, then, might we speed up the computation of parallel analyses?

One reasonably simple thing to do is to parallelize over widely separated observations or batches of observations. Envision two observations on opposite sides of the world. With covariance localization, each observation corrects the background at a mutually exclusive set of grid points. In this case, regardless of whether the two observations are processed serially or simultaneously in parallel, the same analyses will result. This is one simple example of how the computations in the EnKF may be parallelized. Houtekamer and Mitchell (2001) discuss the design and testing of a parallel ensemble Kalman filter exploiting this algorithm.

How else might the work of the ensemble filter be split up among several CPUs? Typically, one of the most computationally expensive steps during the EnKF is the calculation of the $\hat{\mathbf{K}}$, and in particular, the computation of the term $\hat{\mathbf{P}}^{\mathrm{b}}\mathcal{H}^{\mathrm{T}}$. If this could be parallelized, it could result in dramatic cost savings. Following (26), when observations are processed serially, this term is computed from a product of the $n \times m$ matrix of ensemble perturbations and the m-dimensional vector of observation operators applied to the ensemble perturbations. Consider, now, that the n-dimensional state vector is split up over a number of processors. Perhaps the ensemble of northern hemisphere grid points are processed on CPU 1, the southern hemisphere on CPU 2. Say we are to process two observations serially, one just north of the equator and one south of it. For

the first observation, the ensemble of $\left(\mathcal{H}(\mathbf{x}_i^b) - \overline{\mathcal{H}(\mathbf{x}^b)}\right)^T$ from (26) could be computed on CPU 1 and then transferred to processor 2. CPU 1 would then calculate $\hat{\mathbf{P}}^b\mathcal{H}^T$ for the northern hemisphere, CPU 2 for the southern hemisphere. Following that, the update (25) would similarly be split amongst the two CPUs. When the second observation in the southern hemisphere is ready to be processed, the ensemble of $\left(\mathcal{H}(\mathbf{x}_i^b) - \overline{\mathcal{H}(\mathbf{x}^b)}\right)^T$ can be computed on CPU 2 and that data shipped back to CPU 1. The computationally expensive part of the gain calculation and the update are then again computed in parallel. This general method of parallelization could be split up over any arbitrarily large number of CPUs. As long as the operator $\mathcal{H}$ is relatively simple, then the step of computing the $\mathcal{H}(\mathbf{x}_i^b)$ should proceed relatively quickly on one processor (not leaving the other processors idle for long), and the amount of data shipped between processors should be minimal. This technique was demonstrated successfully in Whitaker et al. (2003). See also Keppenne and Rienecker (2002) for another example of parallelization.

c. Variational formulation

The computational expense of most ensemble-based algorithms scales linearly with the number of observations. This may be a large disadvantage if the observations are plentiful, as is to be expected as more and more satellite-based observations are assimilated. The cost of variational methods like 3D-Var (Parrish and Derber 1992) do not scale linearly with the number of observations; practically, doubling the number of observations may increase the costs 10 to 20 % (personal communication, J. Derber). Hence, if the ensemble data assimilation method can be cast in the variational framework, this may provide a computational advantage. This is possible. Starting with the functional in (12), an alternate solution method to (13) is

$$\left(\mathbf{I} + \hat{\mathbf{P}}^b\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)\left(\mathbf{x}_t^a - \mathbf{x}_t^b\right) = \hat{\mathbf{P}}^b\mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{y}_t - \mathbf{H}(\mathbf{x}_t^b)\right) \tag{42}$$

An ensemble-based assimilation method using this formulation was demonstrated in Hamill and Snyder (2000). Though this approach may prove to be less computationally expensive, there are some disadvantages. The adjoint operator of the observation operator will need to be coded, and the technique is iterative in nature, so an effective minimization method and preconditioning

may be needed as well. Covariance localization using this method will be more difficult to implement, since observations are assimilated simultaneously, not serially. Recursive filter methods such as those described in Wu et al. (2002) may be useful for implementing the localization.

7. DEMONSTRATION OF ENSEMBLE-BASED DATA ASSIMILATION

The EnSRF ensemble-based data assimilation methodology discussed in section 4b is now demonstrated in a T62 spectral model. A more in-depth examination of these results is presented in Whitaker et al. (2003), hereafter W03. The purpose of this assimilation experiment was to demonstrate the feasibility of acceptable-quality reanalyses in the decades before radiosonde data was available. Hence, the EnSRF was used to assimilate a network of sparse surface-pressure observations, and the results were compared to 3D-Var. A brief description of the experimental setup is provided here; for more details, see W03.

a. Forecast model

The forecast model was a recent version of the NCEP global medium-range forecast model (MRF), used until 1998. The model was spectral with triangular trunction at wavenumber 62, and it had 28 vertical levels. The model used a sigma (σ) vertical coordinate (Haltiner and Williams 1980). More documentation on this version of the MRF can be found in Wu et al. (1997).

b. Observations

Observations of surface pressure from 2001 were subsampled to resemble the observational surface network of 1915. Observations were assimilated every 6 h, with the bulk of the observations at 0000 UTC and 1200 UTC. A typical 0000 UTC data set had 204 surface pressure observations north of 20^{o} N. latitude (Figure 5). At 0600 and 1800 UTC the density of marine observations was nearly the same, but there are no observations over land. Adjustments were made to the surface-pressure observation when model orography and actual station height differed by less than 1000 m, while stations with station height differences greater than 1000 m were discarded. See W03 for details.

Observation error standard deviations were the same as those used in the NCEP-NCAR reanalysis, 1.6 hPa for ship observations and 1.0 hPa for land stations. Observation errors were assumed to be independent. Observation errors were inflated proportional to the difference between the model orography and actual station elevation.

c. Assimilation methodology

The EnSRF of Whitaker and Hamill (2002) was used as the ensemble-based assimilation methodology. Surface pressure observations were assimilated serially. A horizontal covariance localization of the form in eq. (41) was applied with a correlation length l_c = 2738 km (i.e., grid points farther than 5000 km from the observation are forced to have zero covariance). A vertical covariance localization was also applied. Ensemble-based covariances were directly used below σ = 0.20, not used at all above σ = 0.05, and linearly interpolated in between.

As a crude parameterization of model error, at the beginning of each assimilation cycle, covariances were inflated. Because fewer observations were available in the Southern Hemisphere, the background-error covariances estimated from the ensemble were generally larger. Successive application of a large inflation factor cycle after cycle was thus not warranted, or the ensemble variance would eventually exceed the climatological error variance. Accordingly, the inflation factor r in (39) was set to 1.07 in the Northern Hemisphere and 1.007 in the Southern Hemisphere, with a smooth transition between the two across the equator.

For comparison, a 3D-Var analysis was also computed using just the surface pressure observations. This is the same analysis method used in the NCEP-NCAR "CDAS" (Climate Data Assimilation System) reanalysis, with the following exceptions. First, it was found that better results were obtained with the divergence tendency constraint turned off. Second, because forecast errors are larger with the assimilation of many fewer observations, the standard background-error covariance model was multiplied by a factor of 16, ensuring a closer fit to the available observations. Otherwise, the background-error covariance model was unchanged. It is possible that these static, homogeneous background-error covariances could be tuned further to the sparse observation network, somewhat improving the accuracy of CDAS analyses.

d. Initial conditions and evaluation method

The EnSRF was initialized with an ensemble of random initial conditions drawn from the November NCEP-NCAR reanalysis. The SSI was also initialized with a random reanalysis initial condition. So that the results do not reflect the random choice of initial conditions, the model is allowed to spin up for two weeks starting on 15 November 2001. Unless otherwise noted, the EnSRF results reflect the use of a 100-member ensemble.

All evaluations were done by comparing the analyses against NCEP-NCAR reanalyses, which used several orders of magnitude more observations. Results were evaluated during Dec 2001 at grid points north of 20^{o} N latitude.

e. Results

Figure 4 presents a snapshot of a 500 hPa analysis from the EnSRF as well as from the verifying NCEP-NCAR reanalysis. As shown, the position of long waves and short waves were generally well analyzed, especially in regions with ample observations and downstream of these regions. Short waves in the more data-sparse polar regions and in Siberia were not as well analyzed.

Figure 5 presents a time series of analysis errors from the EnSRF and the CDAS 3D-Var analysis system. The EnSRF was uniformly better than the CDAS, with an average RMS error almost half as small. The error of an EnSRF analysis at 500 hPa was generally about the same magnitude as current 2.5-day forecast errors (W03). Also note in Figure 5 that the spread in the EnSRF was tuned so that on average it is comparable in size to the EnSRF ensemble mean error, as theory suggests it should.

Assimilation experiments were run with ensembles from sizes of 25 to 200 (not shown; see W03). Generally, there was little improvement from the use of ensembles larger than 100 members. The larger ensembles are useful for illustrating the necessity of covariance localization, however. Following Houtekamer and Mitchell (1998), Figure 6 provides a map of sea-level pressure correlations at grid points around the Northern Hemisphere with a grid point in the western

Pacific Ocean on 0000 UTC 14 December 2001. When directly estimated using the 25-member ensemble, correlations for grid points in the region around the observation are positive. The shape of the correlation function was anisotropic, with positive correlations generally limited to a region east of the axis of the cyclone. Background errors for regions in the eastern Pacific and near the Greenwich meridian also appeared to be highly correlated with background errors at the observation location. However, when the correlations are estimated from a 200-member ensemble, it's apparent that these distant correlations in the 25-member ensemble were caused by the limited sample size. The errors in the eastern Pacific and along the Greenwich meridian were not dynamically interconnected with the errors in the western Pacific. When the covariance localization function (Fig. 6c) was applied to the 25-member ensemble, the correlation model (Fig. 6d) more closely resembles that from the larger ensemble.

The EnSRF achieves its improved accuracy over the 3D-Var system in CDAS through the use of flow-dependent background-error covariance statistics. Figure 7 illustrates a snapshot of these covariances in the regions surrounding 5 different observation locations. As shown, the shape of the covariances differed from one observation location to the next, as well as the magnitude of the covariances. For example, the background errors in the region over Siberia and the adjacent Arctic ocean were apparently quite large, yet relatively small over the Atlantic, where ship observations were plentiful. Background-error covariances also had detailed vertical structure. Figure 8 provides cross-covariances between sea-level pressure and geopotential at two levels, 1000 and 500 hPa, for two different surface observation locations. Positive covariances at both locations tilted westward with height. For the observation southwest of the cyclone, the covariances became larger with height, indicating that the surface observation would make a larger correction to the geopotential height at 500 hPa than at 1000 hPa. Conversely, for the observation to the east of the cyclone, the covariances were similar in magnitude. Note that the covariances for the second observation location were much larger, consistent with this location being further from the relatively data-dense land areas, where forecasts were more accurate. Also note that the covariances with observation east of low suggest that the position of the

downstream trough should have been adjusted as well. These covariances were far more complicated in structure than the covariance models used in operational 3D-Var schemes, which typically produce are barotropic, isotropic, and relatively homogeneous. The flow-dependent covariances provide the improvement noted over the simpler 3D-Var.

8. WHERE NEXT?

The field of ensemble-based atmospheric data assimilation is a very new one. To this date, because of the computational expense and the need for a basic understanding of these approaches, most of the experimentation with ensemble assimilation methods have been done in simple models, often using assumptions that are unrealistic for practical numerical weather prediction (for example, the assumption of a perfect model). The results with these simple experiments have been uniformly impressive, indicating that testing in more realistic scenarios is warranted. That is the state of the field in 2003. Several groups, most notably researchers with Environment Canada and in the U.S. National Oceanic and Atmospheric Administration (NOAA), are testing the methodology with real observational data and comparing results against current operational methods.

There still is much to learn about ensemble-based data assimilation methodologies. Parameterizations of model error are just beginning to be explored. The effects of mis-specifying other errors (such as the representativeness errors of observations) on ensemble assimilations is not well understood. The extent to which the underlying assumptions of these filters are met (such as Gaussianity) are not well known. Many practical problems such as ensuring balanced initial conditions may need to be addressed. Further ways to minimize computational expense should be explored. And finally, head-to-head comparisons against 4D-Var in a scenario of identical computing resources should be performed; right now the appeal of ensemble-based methods is more theoretical than evidence-based.

Despite the questions, the potential upside of ensemble assimilation methodologies is large. Ensemble methods are much easier to code and maintain. Ensemble methods cycle the information on background-error covariances, potentially providing a more accurate blending of ob-

servations with the background. Unlike 4D-Var and Kalman filter methods, linearity of error growth is not assumed, and if model error can be parameterized effectively, its effect on the data assimilation can be rationally incorporated.

Because of the potential benefits and the promise of improved objective analyses, expect the literature on ensemble-based data assimilation to grow rapidly in the coming years.

9. ACKNOWLEDGMENTS

Deszo Devenyi (NOAA/FSL), Chris Snyder (NCAR/MMM), Jeff Whitaker (NOAA/CDC), Ryan Torn (U. Washington), and Jim Hansen (MIT) are thanked for their informal reviews of early versions of this manuscript. Three anonymous reviewers provided useful feedback that substantially improved the quality of this manuscript.

This paper was originally prepared for ECMWF's 2002 Predictability Workshop. The many participants who offered interesting questions and constructive criticism at this workshop are thanked as well.

10. APPENDIX: SERIAL PROCESSING OF NON-INDEPENDENT OBSERVATIONS IN THE ENSEMBLE-BASED ASSIMILATION

As previously outlined, if observations have independent errors, then they can be processed serially, which may be of computational advantage. If the observations are not independent, the EnKF update equation (19) can be transformed so that serial processing of observations can occur.

Recall that $\mathbf{y}_t = \mathcal{H}(\mathbf{x}_t^t) + \epsilon$, where $\epsilon \sim N(0, \mathbf{R})$, where $\mathbf{R}$ is no longer assumed to be diagonal. However, $\mathbf{R}$ is symmetric and positive definite, so it has a decomposition of the form $\mathbf{R} = \mathbf{Q}_R \Lambda_R \mathbf{Q}_R^T$, where $\mathbf{Q}_R$ is a unitary matrix with properties that $\mathbf{Q}_R \mathbf{Q}_R^T = \mathbf{I}$ and $\mathbf{Q}_R^T = \mathbf{Q}_R^{-1}$ (here $\mathbf{Q}_R$ does not denote model error). Λ_R is a diagonal matrix of associated eigenvalues.

Let's denote a pseudo-observation $\tilde{\mathbf{y}} = \mathbf{Q}_R^T \mathbf{y}$, or alternately, $\mathbf{y} = \mathbf{Q}_R \tilde{\mathbf{y}}$. Then $\tilde{\mathbf{y}} = \mathbf{Q}_R^T \mathcal{H}(\mathbf{x}_t^t) + \mathbf{Q}_R^T \epsilon$. Hence

$$\left\langle \mathbf{Q}_R^T \epsilon \, (\mathbf{Q}_R^T \epsilon)^T \right\rangle = \mathbf{Q}_R^T \left\langle \epsilon \epsilon^T \right\rangle \mathbf{Q}_R = \Lambda_R. \tag{A20}$$

Define $\widetilde{\mathcal{H}} = \mathbf{Q}_R^T \mathcal{H}$, or equivalently $\mathcal{H} = \mathbf{Q}_R \widetilde{\mathcal{H}}$. Substituting this definition of $\mathcal{H}$ and $\mathbf{y}$ into the EnKF update equation (25), we get

$$\begin{aligned}
\mathbf{x}_i^a &= \mathbf{x}_i^b + \hat{\mathbf{P}}^b \mathcal{H}^T (\mathcal{H} \hat{\mathbf{P}}^b \mathcal{H}^T + \mathbf{R})^{-1} \left(\mathbf{y}_i - \mathcal{H}(\mathbf{x}_i^b) \right) \\
&= \mathbf{x}_i^b + \hat{\mathbf{P}}^b (\mathbf{Q}_R \widetilde{\mathcal{H}})^T (\mathbf{Q}_R \widetilde{\mathcal{H}} \mathbf{P}^b \widetilde{\mathcal{H}}^T \mathbf{Q}_R^T + \mathbf{Q}_R \Lambda_R \mathbf{Q}_R^T)^{-1} \left(\mathbf{Q}_R \tilde{\mathbf{y}}_i - \mathbf{Q}_R \widetilde{\mathcal{H}}(\mathbf{x}_i^b) \right) \\
&= \mathbf{x}_i^b + \hat{\mathbf{P}}^b \widetilde{\mathcal{H}}^T \mathbf{Q}_R^T \mathbf{Q}_R (\widetilde{\mathcal{H}} \hat{\mathbf{P}}^b \widetilde{\mathcal{H}}^T + \Lambda_R)^{-1} \mathbf{Q}_R^T \mathbf{Q}_R \left(\tilde{\mathbf{y}}_i - \widetilde{\mathcal{H}}(\mathbf{x}_i^b) \right) \\
&= \mathbf{x}_i^b + \hat{\mathbf{P}}^b \widetilde{\mathcal{H}}^T (\widetilde{\mathcal{H}} \hat{\mathbf{P}}^b \widetilde{\mathcal{H}}^T + \Lambda_R)^{-1} \left(\tilde{\mathbf{y}} - \widetilde{\mathcal{H}}(\mathbf{x}_i^b) \right).
\end{aligned} \tag{A21}$$

Thus, given a batch of observations with correlated errors and known observation-error covariance matrix $\mathbf{R}$ for these observations, one determines the eigenvectors $\mathbf{Q}_R$ and eigenvalues Λ_R of $\mathbf{R}$, forms the transformed perturbed observations $\tilde{\mathbf{y}}$ and operator $\widetilde{\mathcal{H}}$ and then solves the last line of (A21) can be used to serially process observations. See Kaminsky et al. (1971) for an essentially equivalent algorithm using a Cholesky decomposition.

REFERENCES

Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741-2758.

———, 2001: An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884-2903.

———, 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, in press.

Bellman, R. E., 1961: *Adaptive Control Processes*. Princeton University Press.

Bennett, A. F., B. S. Chua, and L. M. Leslie, 1996: Generalized inversion of a global numerical weather prediction model. *Met. Atmos. Phys.*, **60**, 165-178.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part 1: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420-436.

Blanchet, I., C. Frankignoul, and M. A. Cane, 1997: A comparison of adaptive Kalman filters for a tropical ocean model. *Mon. Wea. Rev.*, **125**, 40-58.

Bouttier, F., 1994: A dynamical estimation of forecast error covariances in an assimilation system. *Mon. Wea. Rev.*, **122**, 2376-2390.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887-2908.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719-1724.

Casella, G. , and R. L. Berger, 1990: *Statistical inference.* Duxbury Press, 650 pp.

Cohn, S. E., and D. F. Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions. *Mon. Wea. Rev.*, **119**, 1757-1785.

———, 1997: An introduction to estimation theory. *J. Meteor. Soc. Jap.*, **75(1B)**, 257-288.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367-1387.

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press. 457 pp.

———, 1992: Estimating model-error covariances for applications to atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 1735-1746.

———, 1993: Estimating observation error statistics for atmospheric data assimilation. *Annales Geophysicae*, **11**, 634-647.

———, 1997: Atmospheric data assimilation. *J. Meteor. Soc. Jap.*, **75(1B)**, 319-329.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128-1145.

Dee, D. P., and R. Todling, 2000: Data assimilation in the presence of forecast bias: the GEOS moisture analysis. *Mon. Wea. Rev.*, **128**, 3268-3282.

Ehrendorfer, M., 1994a: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: theory. *Mon. Wea. Rev.*, **122**, 703-713.

Ehrendorfer, M., 1994b: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part II: applications. *Mon. Wea. Rev.*, **122**, 714-728.

Etherton, B. J., and C. H. Bishop, 2003: Resilience of hybrid ensemble/3D-Var analysis schemes to model error and ensemble covariance error. *Mon. Wea. Rev.*, in review. Available from betherto@rsmas.miami.edu .

Evans, R. E., M. S. J. Harrison, and R. J. Graham, 2000: Joint medium-range ensembles from the Met. Office and ECMWF systems. *Mon. Wea. Rev.*, **128**, 3104-3127.

Evensen, G., 1992: Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.*, **97**, 17905-17924.

———, 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99 (C5)**, 10143-10162.

——— , and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.*, **124**, 85-96.

——— , 2003: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, submitted. Available from Geir.Evensen@nrsc.no

Farrell, B. F., and P. J. Ioannou, 2001: State estimation using a reduced- order Kalman filter. *J. Atmos. Sci.*, **58**, 3666-3680.

Fisher, M., 1998: Development of a simplified Kalman filter *ECMWF Research Department Techical Memorandum 260*. European Centre for Medium-Range Weather Forecasts. 16 pp. Available from Library, ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, England.

Gardiner, C. W., 1985: *Handbook of Stochastic Methods (2nd ed.)*. Springer, 444 pp.

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723-757.

Gauthier, P., P. Courtier, and P. Moll, 1993: Assimilation of simulated lidar data with a Kalman filter. *Mon. Wea. Rev.*, **121**, 1803-1820.

Gelb, A. (ed.), 1974: *Applied optimal estimation.* MIT Press, 374 pp.

Ghil, M., 1989: Meteorological data assimilation for oceanography. Part 1: description and theoretical framework. *Dyn. Atmos. Oceans*, **13**, 171-218.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings - F*, **140**, 107-113.

Haltiner, G. J., and R. T. Williams, 1980: *Numerical Prediction and Dynamic Meteorology*, John Wiley and Sons Press, 477 pp.

Hamill, T. M., and C. Snyder, 2000. A hybrid ensemble Kalman filter / 3d-variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905-2919.

——— , J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background-error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776-2790.

——— , and ——— , 2002: Using improved background-error covariances from an ensemble Kalman filter for adaptive observations. *Mon. Wea. Rev.*, **130**, 1552-1572.

——— , C. Snyder, and J. S. Whitaker, 2003: Ensemble forecasts and the properties of flow-dependent analysis-error covariance singular vectors. *Mon. Wea. Rev.*, **131**, 1741-1758.

Hansen, J. A., 2002: Accounting for model error in ensemble-based state estimation and forecasting. *Mon. Wea. Rev.*, **130**, 2373-2391.

Harrison, M. S. J., T. N. Palmer, D. S. Richardson, and R. Buizza, 1999: Analysis and model dependencies in medium-range ensembles: two transplant case studies. *Quart. J. Roy. Meteor. Soc.*, **125**, 2487-2515.

Hastie, T., R. Tibshirani, and J. Friedman, 2001: *The Elements of Statistical Learning*. Springer, 533 pp.

Heemink, A. W., M. Verlaan, and A. J. Segers, 2001: Variance-reduced ensemble Kalman filtering. *Mon. Wea. Rev.*, **129**, 1718-1728.

Hou, D., E. Kalnay, and K. K. Droegemeier, 2001: Objective verification of the SAMEX-98 ensemble forecasts. *Mon. Wea. Rev.*, **129**, 73-91.

Houtekamer, P. L., 1993: Global and local skill forecasts. *Mon. Wea. Rev.*, **121**, 1834-1846.

——— , L. Lefaivre, and J. Derome, 1996a: The RPN ensemble prediction system. *Proc. ECMWF Seminar on Predictability, Vol II*, Reading, United Kingdom, 121-146. [Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom].

——— , J. Derome, H. Ritchie, and H. L. Mitchell, 1996b: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225-1242.

——— , and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796-811.

——— , and H. L. Mitchell, 1999: Reply to comment on "Data assimilation using an ensemble Kalman filter technique." *Mon. Wea. Rev.*, **127**, 1378-1379.

——— , and ——— , 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123-137.

Jazwinski, A. H., 1970: Stochastic processes and filtering theory. Academic Press, 376 pp.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Transactions of the AMSE- Journal of Basic Engineering.* **82D**, 35-45.

———, and R. S. Bucy, 1961: New results in linear filtering and prediction theory. *Transactions of the AMSE-Journal of Basic Engineering.* **83D**, 95-108.

Kaminsky, P. G., A. E. Bryson, Jr., and S. F. Schmidt, 1971: Discrete square root filtering: a survey of current techniques. *IEE Transactions on Automatic Control*, **AC-16**, 727-736.

Keppenne, C. L., 2000: Data assimilation into a primitive equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 1971-1981.

———, and M. M. Rienecker, 2002: Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.*, **130**, 2951-2965.

Kistler, R., and Coauthors, 2001: The NCEP-NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247-268.

Lacarra, J. F., and O. Talagrand, 1988: Short-range evolution of small perturbations in a barotropic model. *Tellus*, **40A**, 81-95.

Le Dimet, F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, , **38A**, 97-110.

Leith, C. E., 1983: Predictability in theory and practice. Chapter 13 in *Large-scale Dynamical Processes in the Atmosphere.* B. J. Hoskins and R. P. Pearce, Eds., Academic Press, 397 pp.

Lermusiaux, P. F. J., and A. R. Robinson, 1999: Data assimilation via error subspace statistical estimation. Part 1: theory and schemes. *Mon. Wea. Rev.*, **127**, 1385-1407.

———, 2002: On the mapping of multivariate geophysical fields: sensitivities to size, scales, and dynamics. *J. Atmos. Oceanic Technol.*, **19**, 1602-1637.

Li, Z. and I. M. Navon, 2001: Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Quart. J. Roy. Meteor. Soc.*, **127**, 661-683.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177-1194.

——— , 2003: The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, accepted. Available from aclorenc@metoffice.com .

Maybeck, P. S., 1979: *Stochastic models, estimation, and control*. Academic Press, volume 1, chapter 7, 368-409.

Miller, R. N., E. F. Carter, and S. T. Blue, 1999: Data assimilation into nonlinear stochastic models. *Tellus*, **51A**, 167-194.

Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416-433.

——— , and ——— , and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 2791-2808.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73-119.

Ott, E., B. R. Hunt, I. Szunyogh, M. Corazza, E. Kalnay, D.J. Patil, J. A. Yorke, A. V. Zimin, and E. J. Kostelich, 2003: Exploiting local low dimensionality of the atmospheric dynamics for efficient ensemble Kalman filtering. *Mon. Wea. Rev.*, submitted. Available from eo4@umail.umd.edu .

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: a proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279-304.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical Interpolation Analysis System. *Mon. Wea. Rev.*, **120**, 1747-1763.

Penland, C., 2003: A stochastic approach to nonlinear dynamics: a review. *Bull. Amer. Meteor. Soc.*, (Electronic supplement to "Noise out of chaos and why it won't go away, *Bull. Amer. Meteor. Soc.*, , **84**, 921-925).

Pham, D. T., 2001 : Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194-1207.

Potter, J., 1964: W matrix augmentation. *M. I. T. Instrumentation Laboratory Memo SGA 5-64*, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Rabier, F., J.-N. Thepaut, and P. Courtier, 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quart. J. Roy. Meteor. Soc.*, **124**, 1-39.

——— , H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143-1170.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, 2002a: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 103-114.

——— , J. P. Walker, R. D. Koster, and P. R. Houser, 2002b: Extended vs. ensemble Kalman filtering for land data assimilation. *J. Hydrometeorol.*, **3**, 728-740.

——— , and R. D. Koster, 2003: Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. *J. Hydrometeorol.*, accepted. Available from reichle@janus.gsfc.nasa.gov .

Richardson, D. S., 2000: Ensembles using multiple models and analyses. *Quart. J. Roy. Meteor. Soc.*, **127**, 1847-1864.

Sardeshmukh, P. D., C. Penland, and M. Newman, 2001: Rossby waves in a stochastically fluctuating medium. In *Stochastic Climate Models*, ed. P. Imkeller and J.-S. von Storch, Progress in Probability, Birkhaueser, Basel, pp. 369-384.

Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663-1677.

Talagrand, O., 1997: Assimilation of observations, an introduction. *J. Meteor. Soc. Jap.*, **75(1B)**, 191-209.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.,* **131** , 1485-1490.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.,* **74**, 2317-2330.

——— , and ——— , 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.,* **12**, 3297-3319.

van Leeuwen, P.J., 1999: Comment on "Data assimilation using an ensemble Kalman filter technique." *Mon. Wea. Rev.,* **127**, 1374-1377.

Verlaan, M. and A. W. Heemink, 2001: Nonlinearity in data assimilation applications. A practical method for analysis. *Mon. Wea. Rev.,* **129**, 1578-1589.

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.,* **60**, 1140-1158.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed obsevations. *Mon. Wea. Rev.,* **130**, 1913-1924.

——— , G. P. Compo, X. Wei, and T. M. Hamill, 2003: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, submitted. Available from jeffrey.s.whitaker@noaa.gov .

Wu, W.-S., M. Iridell, S. Saha, and P. Caplan, 1997: Changes to the 1997 NCEP operational MRF model analysis/forecast system. *Technical Procedures Bulletin 443*, NCEP, National Weather Service, Office of Meteorology, Programs and Plans Division, Silver Spring MD, 20910. Available from http://www.nws.noaa.gov/om/tpb/indexb.htm .

——— , R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.,* **130**, 2905-2916.

Zhang, F., C. Snyder, and J. Sun, 2003: Impact of initial estimate and observations on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, in review. Available at http://www.met.tamu.edu/personnel/faculty/fzhang/PAPERS/ZhangSnyderSun2003MWR.pdf

Ziehmann, C, 2000: Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models. *Tellus*, **52a**, 280-299.

FIGURE CAPTIONS

Figure 1. Example of Bayesian data assimilation update. Here the model state is two dimensional and a single observation is assimilated. This observation measures the same variable as the first component of the model state. (a) Probability density for prior joint and marginal distributions (solid) and sample observation distribution (dashed). The three contours enclose 75 %, 50 %, and 25 % of the probability density, respectively. (b) Probability density for posterior distributions. Contours levels set as in (a).

Figure 2. Illustration of the EnKF and EnSRF with a two-dimensional state variable and observations observing the same as $x^b_{(1)}$. (a) Random samples (black dots) from the probability distribution in (1), and the original prior pdf, contoured in red. Implied bivariate normal probability background distribution estimated from the sample ensemble contoured in black, and the observation sampling distribution (dashed). Solid vertical lines along abscissa denote individual perturbed observations sampled from this distribution. The one large black dot and the perturbed observation marked with a star denote the sample discussed in the text. (b) Analyzed samples from the EnKF assimilation scheme (dots), the implied analysis-error bivariate normal distribution from this sample (solid black contours), and the true posterior pdf from Fig. 1 (red). (c) Analyzed samples from EnSRF (dots), implied bivariate normal pdf (solid black contours) and the true posterior pdf (red). In each panel, the three contours enclose 75 %, 50 %, and 25 % of the probability density, respectively.

Figure 3. Illustration of covariance localization in domain with 100-dimensional state vector. (a) True covariances model for grid points 1, 21, 41, 61, and 81 (solid lines). Estimate of covariances from 50-member ensemble (dashed). (b) As in (a), but after application of covariance localization with correlation function with correlation length of 15 grid points.

Figure 4. 500 hPa geopotential height analysis for 0000 UTC 14 Dec 2001 (contour interval 50 m). CDAS analysis, using all available observations is shown on the left. EnSRF analysis

using simulated 1915 surface pressure observation network is shown on right. Black dots indicate locations of surface pressure observations used in the EnSRF analysis at this time.

Figure 5. Time series of RMS analysis errors from EnSRF (black line) and CDAS (red line) and EnSRF spread (green line). (a) Errors of mean sea-level pressure (hPa), (b) errors of 500 hPa height.

Figure 6. Illustration of covariance localization. (a) Correlations of sea-level pressure directly estimated from 25-member ensemble with pressure at a point in the western Pacific (colors). Solid lines denote ensemble mean background sea-level pressure contoured every 8 hPa. (b) As in (a), but using 200-member ensemble. (c) Localization function, (d) Correlation estimate from 25-member ensemble after application of covariance localization.

Figure 7. Background-error covariances (colors) of sea-level pressure in the vicinity of five selected observation locations, denoted by dots. Covariance magnitudes are normalized by the largest covariance magnitude on the plot. Solid lines denote ensemble mean background sea-level pressure contoured every 8 hPa.

Figure 8. Cross-covariances of mean sea-level pressure and geopotential height for two selected pressure observation locations. (a) 1000 hPa geopotential height (units of m) and cross-covariance of sea-level pressure (location marked with dot) with 1000 hPa height (units hPa$\times$m). (b) As in (a), but for 500 hPa height. (c) As in (a), but for the second observation location. (d) As in (b), but for the second observation location.

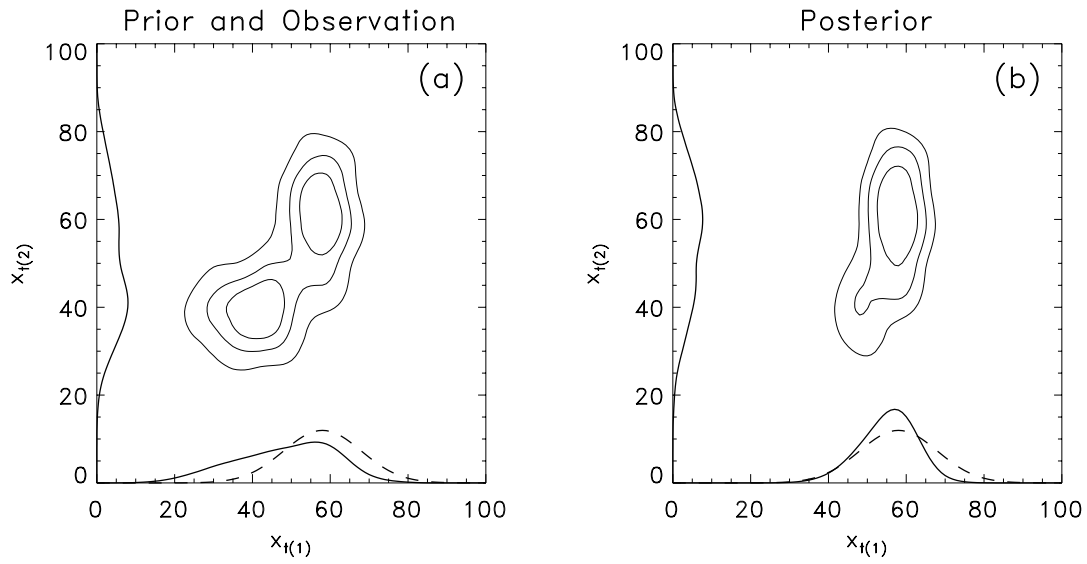

Figure 1. Example of Bayesian data assimilation update. Here the model state is two dimensional and a single observation is assimilated. This observation measures the same variable as the first component of the model state. (a) Probability density for prior joint and marginal distributions (solid) and sample observation distribution (dashed). The three contours enclose 75 %, 50 %, and 25 % of the probability density, respectively. (b) Probability density for posterior distributions. Contours levels set as in (a).

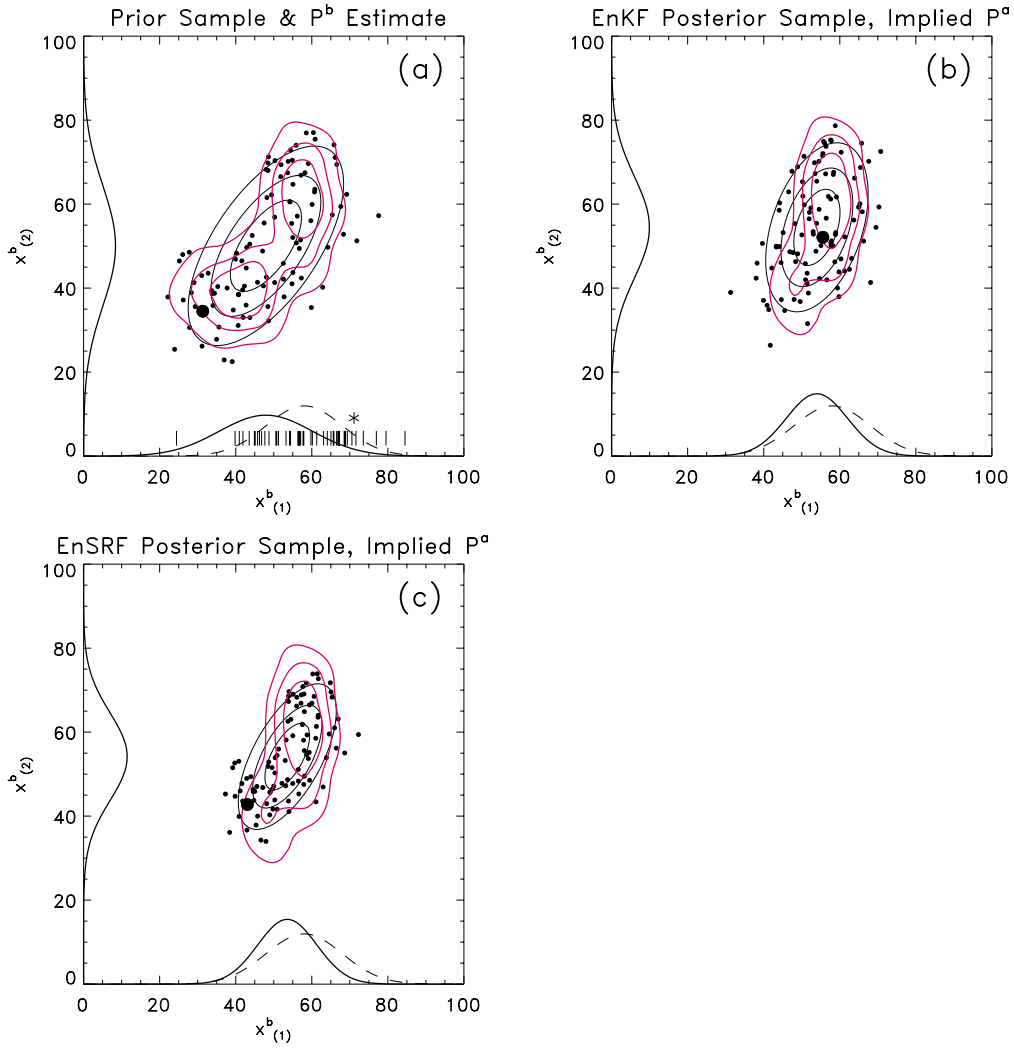

Figure 2. Illustration of the EnKF and EnSRF with a two-dimensional state variable and observations observing the same as $x^b_{(1)}$. (a) Random samples (black dots) from the probability distribution in (1), and the original prior pdf, contoured in red. Implied bivariate normal probability background distribution estimated from the sample ensemble contoured in black, and the observation sampling distribution (dashed). Solid vertical lines along abscissa denote individual perturbed observations sampled from this distribution. The one large black dot and the perturbed observation marked with a star denote the sample discussed in the text. (b) Analyzed samples from the EnKF assimilation scheme (dots), the implied analysis-error bivariate normal distribution from this sample (solid black contours), and the true posterior pdf from Fig. 1 (red). (c) Analyzed samples from EnSRF (dots), implied bivariate normal pdf (solid black contours) and the true posterior pdf (red). In each panel, the three contours enclose 75 %, 50 %, and 25 % of the probability density, respectively.

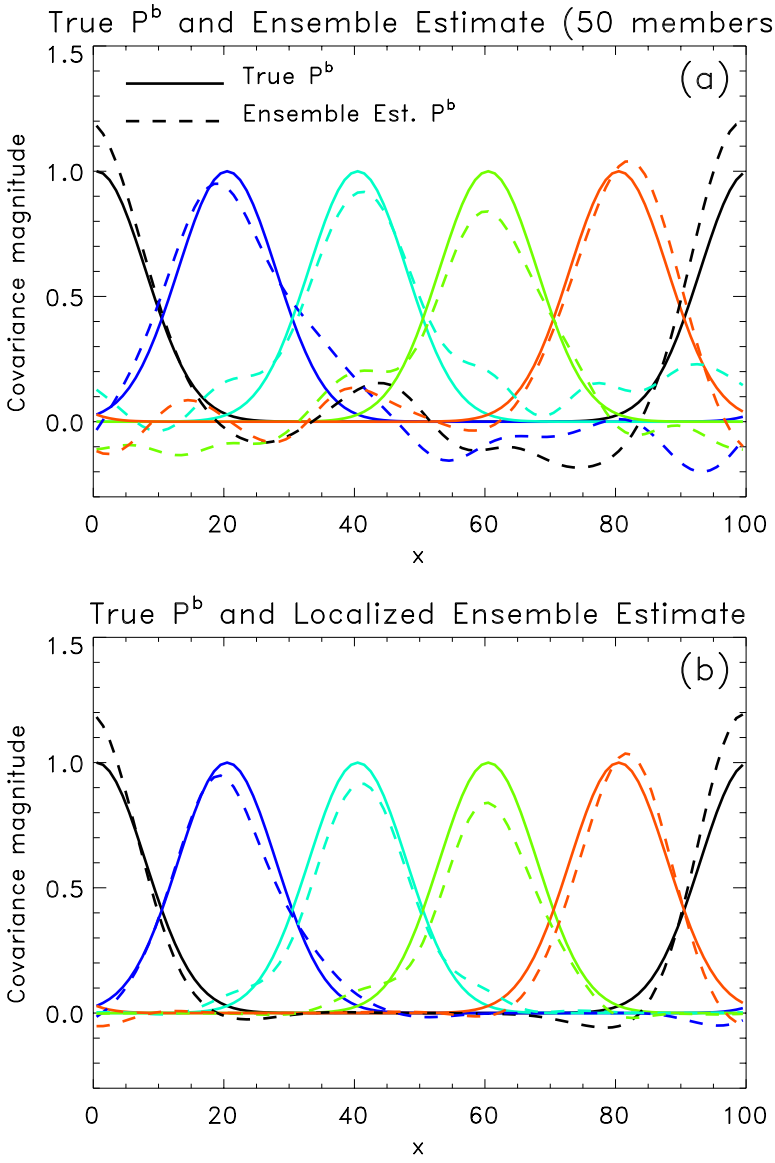

Figure 3. Illustration of covariance localization in domain with 100-dimensional state vector. (a) True covariances model for grid points 1, 21, 41, 61, and 81 (solid lines). Estimate of covariances from 50-member ensemble (dashed). (b) As in (a), but after application of covariance localization with correlation function with correlation length of 15 grid points.

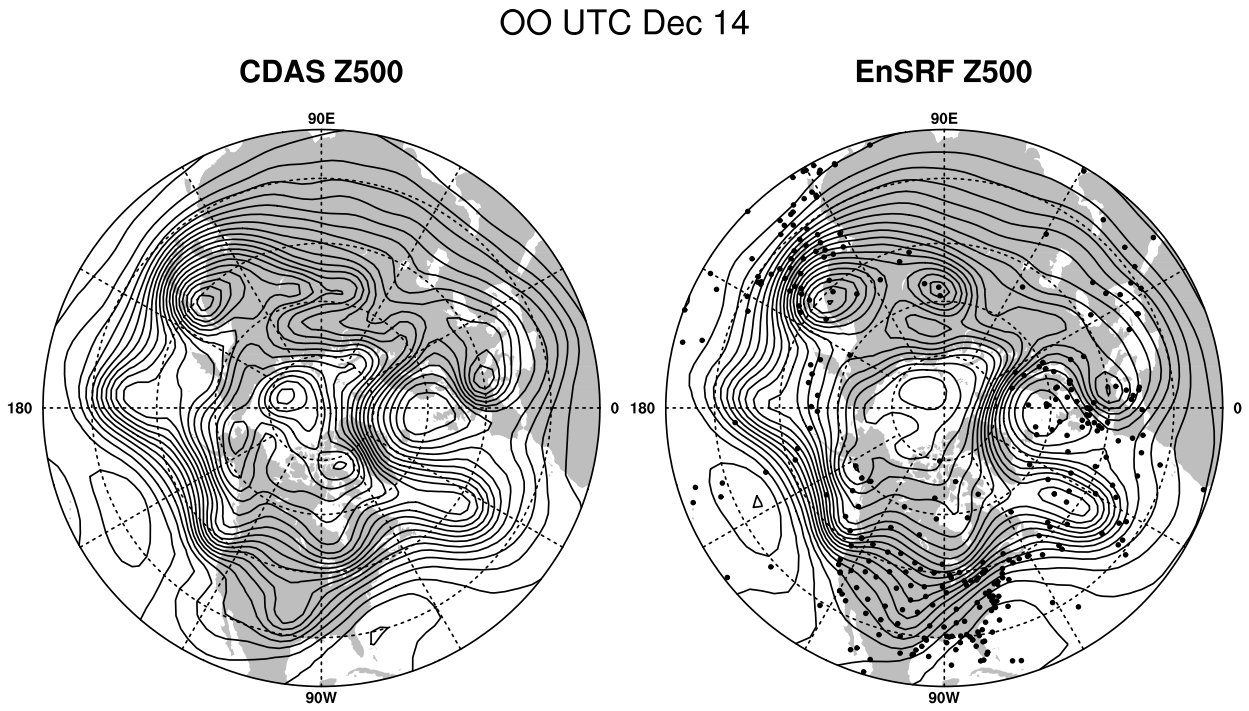

Figure 4. 500 hPa geopotential height analysis for 0000 UTC 14 Dec 2001 (contour interval 50 m). CDAS analysis, using all available observations is shown on the left. EnSRF analysis using simulated 1915 surface pressure observation network is shown on right. Black dots indicate locations of surface pressure observations used in the EnSRF analysis at this time.

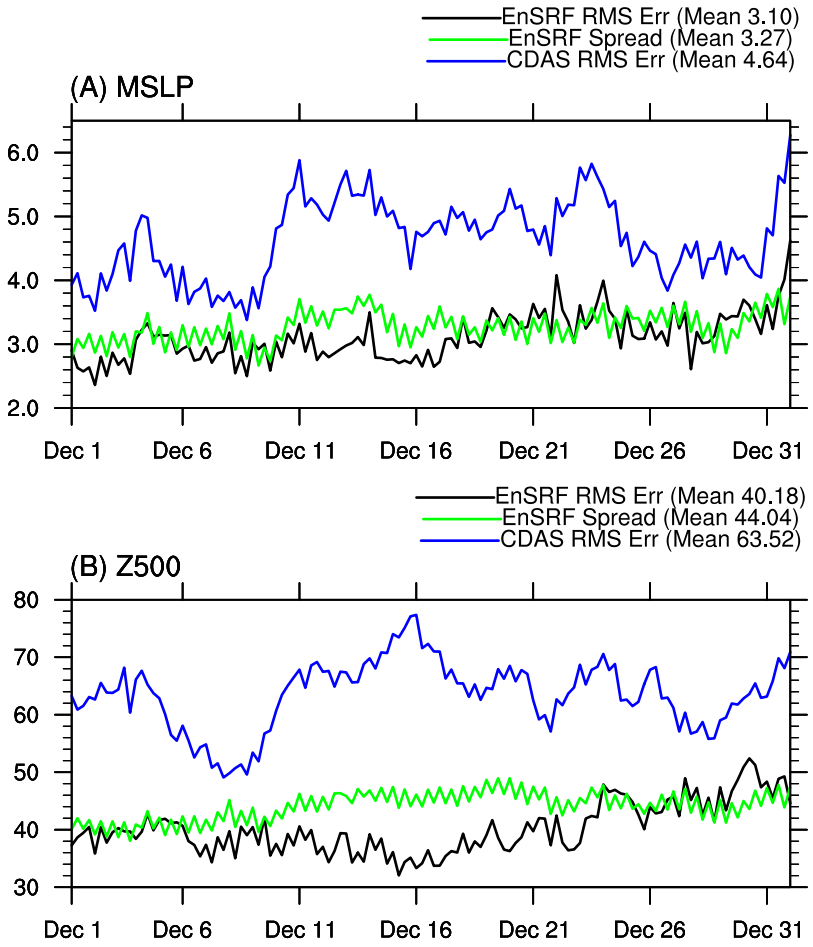

Figure 5. Time series of RMS analysis errors from EnSRF (black line) and CDAS (red line) and EnSRF spread (green line). (a) Errors of mean sea-level pressure (hPa), (b) errors of 500 hPa height.

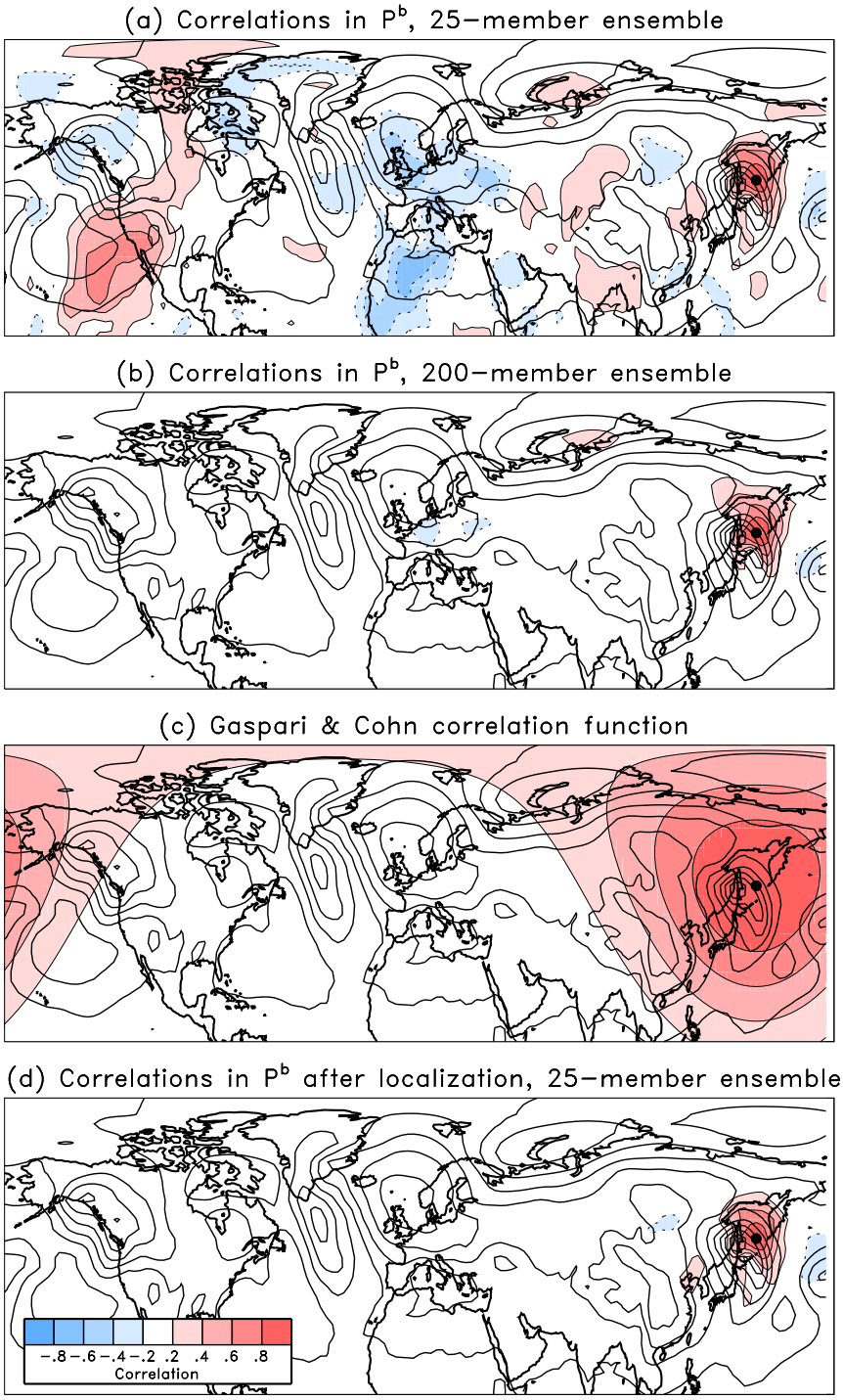

Figure 6. Illustration of covariance localization. (a) Correlations of sea-level pressure directly estimated from 25-member ensemble with pressure at a point in the western Pacific (colors). Solid lines denote ensemble mean background sea-level pressure contoured every 8 hPa. (b) As in (a), but using 200-member ensemble. (c) Localization function, (d) Correlation estimate from 25-member ensemble after application of covariance localization.

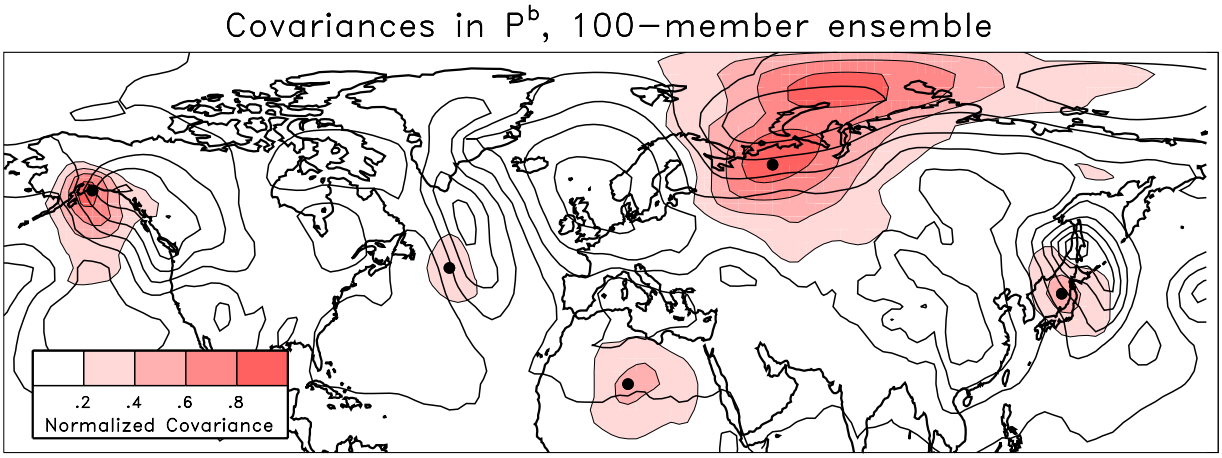

Figure 7. Background-error covariances (colors) of sea-level pressure in the vicinity of five selected observation locations, denoted by dots. Covariance magnitudes are normalized by the largest covariance magnitude on the plot. Solid lines denote ensemble mean background sea-level pressure contoured every 8 hPa.

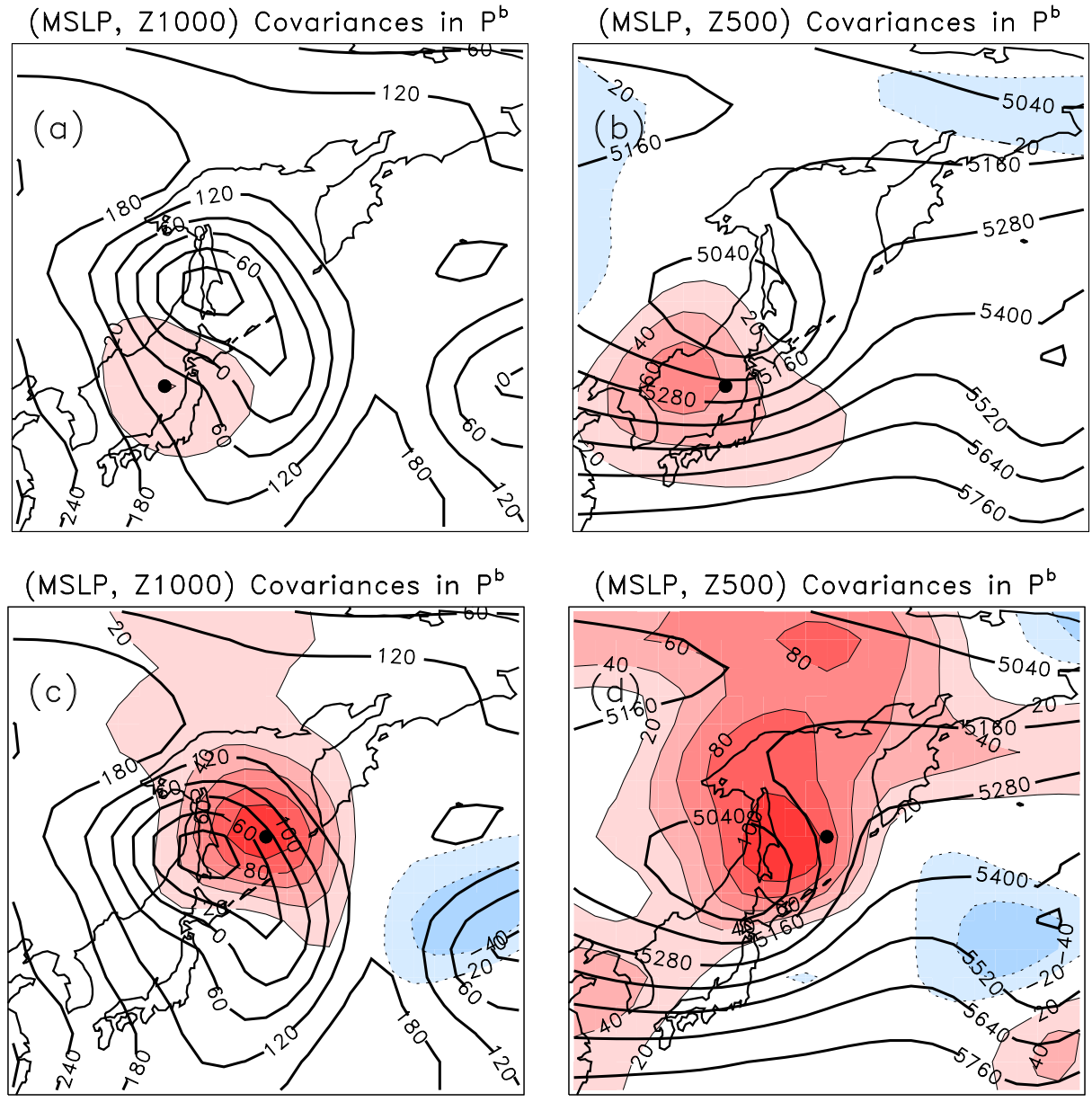

Figure 8. Cross-covariances of mean sea-level pressure and geopotential height for two selected pressure observation locations. (a) 1000 hPa geopotential height (units of m) and cross-covariance of sea-level pressure (location marked with dot) with 1000 hPa height (units hPa×m). (b) As in (a), but for 500 hPa height. (c) As in (a), but for the second observation location. (d) As in (b), but for the second observation location.